*electronics*

MDPI

*Review*

# Survey on Q-Learning-Based Position-Aware Routing Protocols in Flying Ad Hoc Networks

**Muhammad Morshed Alam and Sangman Moh ***

Department of Computer Engineering, Chosun University, 309 Pilmun-daero, Dong-gu, Gwangju 61452, Korea; morshed@chosun.kr

* Correspondence: smmoh@chosun.ac.kr; Tel.: +82-62-230-6032

**Abstract:** A flying ad hoc network (FANETs), also known as a swarm of unmanned aerial vehicles (UAVs), can be deployed in a wide range of applications including surveillance, monitoring, and emergency communications. UAVs must perform real-time communication among themselves and the base station via an efficient routing protocol. However, designing an efficient multihop routing protocol for FANETs is challenging due to high mobility, dynamic topology, limited energy, and short transmission range. Recently, owing to the advantages of multi-objective optimization, Q-learning (QL)-based position-aware routing protocols have improved the performance of routing in FANETs. In his article, we provide a comprehensive review of existing QL-based position-aware routing protocols for FANETs. We rigorously address dynamic topology, mobility models, and the relationship between QL and routing in FANETs, and extensively review the existing QL-based position-aware routing protocols along with their advantages and limitations. Then, we compare the reviewed protocols qualitatively in terms of operational features, characteristics, and performance metrics. We also discuss important open issues and research challenges with potential research directions.

**Keywords:** FANET; UAV network; routing protocol; position-aware routing; multi-objective optimization; reinforcement learning; Q-learning

check for updates

**Citation:** Alam, M.M.; Moh, S. Survey on Q-Learning-Based Position-Aware Routing Protocols in Flying Ad Hoc Networks. *Electronics* **2022**, *11*, 1099. https://doi.org/10.3390/electronics11071099

Academic Editor: Juan M. Corchado

Received: 21 February 2022
Accepted: 29 March 2022
Published: 30 March 2022

## 1. Introduction

In recent years, unmanned aerial vehicle (UAV) networks have received considerable interest in both industrial and academic research owing to their potential applications in surveillance [1], computationally intensive tasks for low-powered Internet-of-Things (IoT) devices [2], wildfire monitoring [3], agricultural surveying, and aerial base stations (ABSs) [4]. UAV swarm networks can efficiently accomplish complex tasks through autonomous collaboration. This is primarily because of the significant development of enabling technologies such as sensors, batteries, formation control, localization based on the global positioning system (GPS), and range-free and range-based localization in GPS-denied environments [5]. Low-altitude UAVs, commonly known as drones, have exhibited considerable promise in a wide variety of applications designed to mitigate disease outbreaks, such as the current COVID-19 pandemic. The UAV swarm can be deployed to perform crowd surveillance, provide public announcements to enforce social distancing, spray disinfectants in contaminated areas, and transport medical supplies to remote areas. In addition, UAVs equipped with thermal cameras can capture large-scale temperature measurements in crowded places [6]. Owing to the flexibility in mobility adjustment, UAVs are also used as a mobile power beacon broadcaster to optimally transfer energy to the low-powered IoT devices [7].

A UAV swarm can collaboratively sense a particular mission area to capture continuous video and three-dimensional (3D) light detection and ranging mapping (LiDAR) images using onboard sensors. The data collected by each UAV must be transmitted to a base station (BS) with a minimum delay and energy-efficient routing for further processing.

This multi-UAV collaborative network, also known as flying ad hoc networks (FANETs), has several key challenges such as a dynamic time-varying topology owing to the mobility of UAVs in 3D space, node density representing the trade-off between sensing coverage and quality of service (QoS) in communication [8], limited transmission range, and limited onboard energy. Although UAV swarms have a wide range of applications, they should generally involve cooperative coordination to achieve mission objectives. Simultaneously, they should consider all constraint resources to optimize the overall network performance to ensure the desired signal-to-interference-plus-noise ratio (SINR) in the aerial links. Therefore, proper collaboration among UAVs is necessary to achieve a balance between mission performance, such as maximizing the coverage to the mission area or ground users (GUs), and communication performance to achieve a desirable connectivity rate with minimal delay in the aerial links.

To achieve coverage efficiency, UAVs may frequently fly away from each other's transmission ranges, which causes frequent link breakages. Frequent link breakages during data transmission may increase the number of retransmissions that consume more energy, increase the number of collisions in the medium access control (MAC) layer, and consume unnecessary bandwidth of wireless links. To maintain a better SINR, UAVs should not frequently fly away from each other's transmission ranges. Simultaneously, UAVs should maintain a certain separation distance to avoid physical collision and maximize sensing coverage toward the ground terminal. Therefore, UAVs in the swarm should maintain a strong neighbor relationship and optimal node density to maintain a stable topology for each time interval via proper topology control by following the three principles of flocking: separation, cohesion, and alignment [9]. The separation rule aids UAVs in maintaining a particular separation distance to avoid inter-UAV collisions and maximize coverage. The cohesion rules define the mobility of each UAV in the swarm to be attracted to the average centroid of the neighboring UAV positions. This aids the UAV swarm in staying close to each other to avoid frequent link breakages. Finally, the alignment rule requires that each UAV adopt a velocity direction according to its neighboring UAVs. Similarly, in complex environments, UAVs can sense the presence of external obstacles and generate external obstacle avoidance rules. Each flocking rule generates one motion component, and the sum of these motion components determines the optimal mobility of UAVs [10]. The adaptive adjustment of the weight of each rule can generate a smoother and more energy-efficient trajectory for UAVs [11].

Topology control in a FANET is a mechanism of coordinating and optimizing the mobility of UAVs (position, velocity, and acceleration) according to the transmission range of each UAV, which can produce a network with optimal transmission power and node density. Topology control reduces energy consumption, improves connectivity, and facilitates wider coverage with desirable throughput by controlling UAV mobility and transmission power [12]. Compared with centralized control, the distributed control of the UAV swarm enhances stability and routing efficiency by maintaining neighbor intimacy utilizing one- or two-hop neighbor information. Centralized control provides less scalability and causes the possibility of a single point of failure. It also consumes a high bandwidth because each UAV must transmit its mobility information to a central node to update the topology. However, considerable theoretical disputes occur regarding UAV swarm control based on partial and relative local knowledge without the intervention of a central controller. Recently, significant developments have been achieved for the distributed control of UAV swarms, in which local optima are avoided by utilizing several metaheuristics [13] and game-theory-based optimization [14,15]. The major limitation of metaheuristics and game theory is their high computational complexity with increased nodes and topology size. Therefore, to satisfy the scalability requirement and solve existing challenges such as delay, energy constraints, and mobility, an intelligent alternative approach that can perform multi-objective optimization adaptively in highly dynamic FANETs is required. Distributed algorithms such as the artificial potential field (APF) [12,16], virtual force [10,17], and boid flocking [9] have attracted the interest of researchers in constructing topology in FANETs by producing optimal mobility of UAVs in FANETs. These algorithms are highly efficient at performing external obstacle avoidance [11], generating a smooth trajectory for UAVs [10], and maximizing coverage under connectivity constraints [18,19]. The adaptive adjustments of their attractive and repulsive force fields also support the maintenance of an optimal node density in FANETs to minimize interference in aerial links [12,20].

In FANETs, link quality depends on several parameters, such as inter-UAV distance, node density, SINR, delay, relative mobility, and residual energy of relaying UAVs. The optimal node density and link SINR can be achieved by jointly optimizing the UAV mobility (position, velocity, and acceleration) and transmitting power according to the inter-UAV distance by adopting a topology control technique [12,21]. The link delay includes MAC-layer channel access, queuing, propagation, processing, and transmission delays. The optimal resource allocation in resource-constrained FANETs, such as physical-layer UAV transmission power, MAC-layer time slots, or frequency resources, can significantly improve the SINR level in aerial links. Thus, this sequentially improves the network-layer performance (relay selection) as they are highly coupled.

Owing to the above advantages, researchers have jointly considered the MAC layer delay, link SINR, relative mobility, position progress to the destination, and residual energy of neighboring UAVs, to design a multi-objective reward function in reinforcement learning (RL)-based algorithms [22–24]. RL is an area of machine learning concerned with how intelligent agents ought to take an action from a specific state by interacting with an environment to maximize rewards. Through the iterative state transitions, an agent learns how to choose an optimal action. Thus, RL-based action can be formulated as a Markov decision process (MDP) tuple consisting of state, action, and reward. The state represents the consequences that an agent faces in a dynamic environment by taking actions according to the learning policy. Through sequential action and utilizing previous experience, RL agents can make wiser decisions to reach a common objective. In communication theory, RL is applied in many scenarios such as channel modeling, resource allocation, and security [25,26]. Recently, RL has been widely used in FANETs to design the smooth collaborative trajectory planning for UAV swarms with collision avoidance [27], and routing protocol design [23]. Q-learning (QL) is a model-free value-based off-policy RL approach, which can obtain an instant optimal policy based on historic experiences even without prior information of the environment or even without the intervention of any central controller [28]. Here, each agent makes an optimal decision based on its neighbor state information, which can be treated as partial MDP (PMDP).

According to our earlier discussion in this section, we can say that the topology controller (formation controller) iteratively updates the mobility of each UAV within a swarm by using the mobility information of its one-hop neighbors. Additionally, the output of the topology controller decides the topology of the UAV swarm by predicting the present and future mobility information for each UAV (acceleration, velocity, position, and flying direction) [21]. Thus, we can say that relative trajectory knowledge given by the formation controller and link stability is highly coupled [29]. It can ensure stable connectivity between UAVs during flocking. The formation controller updates the mobility information for each UAV in the next timeslot based on the mobility information in the current timeslot, which indicates the similarity with the Markov property. This is because the Markov property states that the next states of the process depend only on the current state of the process. As a result, QL-based PMDP formulation can be adopted to make routing decisions to find the most stable path in FANETs. Owing to this relationship, researchers have used the QL technique to select the optimal relay nodes for forwarding data in FANETs by designing a multi-objective reward function. Because the reward function reinforces the action policy of an RL agent and accelerates the algorithm convergence for optimal decision making, a good reward function considering multiple objectives (delay, relay node energy, and distance progress toward the destination node) gives better routing performance in FANETs. Consequently, this joint consideration of multiple objectives significantly improves the packet delivery ratio (PDR), throughput, end-to-end delay, and balances the energy consumption in FANETs. Considering the high mobility, constraint energy, and memory resources of UAVs, the QL method is more suitable for FANET routing decision making than deep reinforcement learning because it is computationally more expensive and requires a large memory to store training samples and a history of action–reward pairs. The relationship between QL and position-based forwarding is discussed further in Section 3.

In FANETs, UAVs can utilize the GPS to localize them in global coordinates. In a GPS-denied environment, UAVs can use range-free and range-based cooperative localization techniques to identify self-location and the location of neighbors [14,19]. Consequently, position-based routing can be effectively used in dynamic FANETs. In this article, we only extensively survey existing QL-based position-aware routing protocols for FANETs. We also discuss open issues and challenges and their potential research directions related to QL-based position-aware routing in FANETs.

In the following subsections, we address related studies and summarize the contributions of this study.

### *1.1. Related Studies*

In this subsection, we discuss recent survey papers related to FANET routing protocols and the limitations of existing FANET routing protocols. Additionally, we discuss the motivation for our research and the key contributions of this survey paper.

According to previous studies [30–32], the routing protocols in FANETs are classified as topology- and position-based. Topology-based routing protocols can be further classified as proactive, reactive, and hybrid routing protocols. Proactive routing protocols produce a large overhead to maintain the updated routing table for a dynamic topology. Thus, they consume higher bandwidth and energy, which is not suitable for resource-constrained FANETs. Additionally, they exhibit a slow reaction to a highly dynamic topology, which causes delays, routing loops, and blind paths [23]. A loop-free property is essential for dynamic FANETs to prevent data packets from being continually routed through similar nodes or paths. Blind path challenges occur in FANETs when the neighboring UAVs leave the transmission range of the corresponding source UAV within the intermediate time of the topology update because of several reasons such as sudden changes in relative mobility, requirements for energy replenishment, and UAV failure [23]. Additionally, FANETs may encounter frequent link breakages if the selected relay UAV leaves the transmission range of the corresponding source UAV during data transmission. Both the blind path and link breakage phenomena produce high retransmissions, delays, and energy consumption in FANETs.

In [33], the authors studied the optimized link-state routing protocol (OLSR), which encounters higher overheads and routing loops and has a slow reaction in highly dynamic networks. Similarly, in [34], the authors studied the destination sequenced distance vector (DSDV), which consumes a large portion of the network bandwidth and provides a very high overhead owing to periodic updates in FANETs. Reactive routing protocols result in higher latency and delays owing to the on-demand route-discovery process. Additionally, in large-scale FANETs, the network overhead increases for reactive routing owing to an increase in the header size of the routing table [35]. In [32], the authors reported that dynamic source routing (DSR) provides a comparatively lower overhead at the cost of delays in route discovery. However, for large-scale FANETs, DSR routing encounters an extremely high overhead owing to an increase in the routing discovery table header [35]. Similarly, ad hoc on-demand distance vector (AODV) routing encounters route failures, higher delays, and higher bandwidth consumption in large-scale FANETs [34]. Hybrid routing protocols encounter higher computational complexities and overhead owing to the complex clustering, cluster-head selection, and cluster maintenance processes [3]. Therefore, all these traditional topology-aware proactive, reactive, and hybrid routing protocols encounter several limitations in highly dynamic FANETs owing to the high control overhead and large delays in neighbor and path discovery [36]. Additionally, they do not support adaptability to the dynamic topology to discover the efficient routing path autonomously.

In position-based routing, each UAV node utilizes the GPS for localization. In addition, UAVs can use range-free and range-based cooperative localization in a GPS-denied environment. Position-based routing protocols utilize local knowledge, often one- or two-hop information, to make routing decisions. UAVs make forwarding decisions based on their current position, the position of the destination, and the position of their neighbors. In [37,38], the authors studied several position-based routing protocols in FANETs by classifying them into the two categories of single-path and multipath strategies. Under the single-path strategy, they reviewed deterministic progress-based, randomized progress-based, and hybrid position-based routing protocols. Deterministic progress-based routing protocols have several relay node-selection strategies, including greedy forwarding, compass forwarding, and most forwarding [37]. Multipath strategies include restricted direction flooding, random directional flooding, and classic flooding of data packets [37].

According to the aforementioned study, considering the dynamism in network topology in the 3D space, inter-UAV collision, high overhead, and delay, position-based routing protocols are attracting the interest of researchers. However, position-based routing protocols encounter several challenges in FANETs, such as maintaining the link quality [39], controlling the hello interval to predict up-to-date topology [21], localization errors, blind paths, the presence of routing loops, and energy holes [23]. Additionally, to prolong the lifetime of a FANET, it is necessary to achieve a proper load balance in terms of energy and delay while determining the optimal routing path [22]. Tracing the shortest routing path may be initially beneficial, but it cannot be an optimal routing path as it depletes the energy of a few selected UAVs, and the shortest paths can be extremely congested by traffic over time [23]. It also creates energy holes in FANETs because selecting the shortest path always drains the energy of a few selected UAVs. Greedy forwarding cannot ensure optimal performance in terms of energy consumption, delay, and link quality, as it always seeks progress in the transmission distance toward the destination. Additionally, owing to the selection of relay nodes at the edge of the transmission range of the source node, greedy forwarding encounters blind path and link-breakage problems. The compass and most forward techniques have higher possibilities of trapping in routing loops and local minimum [37]. The term local minimum (routing holes) in position-based routing is defined as the selected relay UAV with no further neighbors to relay toward the target destination node. Flooding techniques in multipath forwarding produce excessive overhead, high MAC layer contention, high bandwidth, and energy consumption.

UAVs in a swarm exchange hello packets to update their position coordinates and residual energy. The low hello-interval provides better positioning accuracy but simultaneously increases the control overhead cost [21]. Consequently, an adaptive strategy is required to determine the optimal hello interval in FANETs to optimize the control overhead cost and predict the updated network topology. To address the above challenges in position-based forwarding, researchers have proposed intelligent decision-making algorithms in dynamic FANETs utilizing the QL technology incorporated with position-based routing protocols. QL-based routing protocols can perform multi-objective optimization; i.e., delay and energy are minimized by leveraging the PMDP for predicting the dynamic topology with the aid of a topology controller. Recently, a significant amount of research has been conducted to enhance the performance of position-based forwarding techniques by integrating them with QL. The QL model can be trained to identify a link that is trapped in the local minimum in position-based forwarding by providing a minimum reward to the relay nodes for taking a bad action. Additionally, in [22,24], the designed QL-model allocates minimum reward to the relay UAVs that do not send the acknowledgment to the corresponding source UAV by considering the failure state of the selected relay nodes. In FANETs, the UAV failure state might happen owing to the hardware failure, the depletion of UAV energy, and the environmental dynamism encountered by the UAVs (external obstacles). Thus, a robust self-healing topology controller is required to reestablish the swam topology without creating any partition in topology, while performing the mission in a complex dynamic environment [11,17].

So far, the comprehensive review articles on position-based routing protocols in FANETs discuss the different relay UAV selection mechanisms by considering the path selection strategy, bio-inspired swarming, and topology-based routing [37,38]. The QL-based routing protocols incorporated with position-based forwarding techniques are new research trends for FANETs, which are not covered by the existing survey works. Motivated by this, we surveyed all the recently proposed QL-based position-aware routing protocols in FANETs, most of which are published in reputed journals. In our earlier study [40], the relationship between QL and routing was reviewed, and seven QL-based routing protocols for FANETs were discussed and compared qualitatively, primarily focusing on key features and performance challenges. However, in this review article, we define the FANETs, their components with functionalities, and dynamic FANET topology. Then, we extensively study the realistic mobility models in FANETs according to their applications and define the QL and its relationship with dynamic FANETs. Additionally, we comprehensively review each protocol with its advantages and limitations, and then perform a comparative study considering important performance metrics for FANETs. According to our comparative study, we find key open issues and discuss their potential research directions.

According to the above discussion, related review articles on FANET routing protocols are summarized in Table 1 to indicate our key contributions.

**Table 1.** Comparative summary of survey works and their focused points on routing in FANETs.

| Ref. | Year | Key Focused Points |
|---|---|---|
| [32] | 2019 | Reviewed the routing protocols in FANETs by classifying them as topology-based, position-based, and hierarchical routing. |
| [30] | 2019 | Reviewed UAV types, mobility models, and routing protocols in FANETs by classifying them as topology-based, position-based, and delay-tolerant-based categories. |
| [31] | 2020 | Reviewed UAV mobility models, application scenarios of UAV networks, and routing protocols by classifying them as topology-based, position-based, hybrid, and bio-inspired routing. |
| [37] | 2018 | Reviewed the only position-based routing protocols in FANETs by classifying them based on the path selection strategy. |
| [38] | 2017 | Reviewed the only position-based routing protocols in FANETs based on the three categories of topology-based, position-based, and swarm-based routing. |
| This survey | 2022 | • Reviews the mobility model of the UAV swarm networks based on the realistic application scenarios in FANETs<br>• Defines adaptive QL and its relationship with routing in dynamic FANETs<br>• Reviews the QL-based position-aware routing protocols in FANETs with advantages and limitations<br>• Summarizes performance enhancements criteria (in the Lessons Learned section) according to our comparative study<br>• Addresses the key open challenges and potential research directions |

The limitations of topology-based proactive, reactive, and hybrid routing protocols are summarized in Table 2.

**Table 2.** Summary of topology-based routing protocols and their limitations in FANETs.

| Protocol Type | Limitations to Adopt in FANETs |
|---|---|
| Proactive | High control overhead and bandwidth consumption to maintain an updated neighbor table, and slow reaction to rapid topology changes. |
| OLSR [33] | High control overhead, routing loop, and link breakage. |
| DSDV [34] | Requires periodic updates, high bandwidth, and control overhead |
| Reactive | Higher delay and latency in routing discovery, and no link quality assessment |
| DSR [36] | Produces high overhead during route discovery in large-scale networks, and high delays |
| AODV [34] | Higher delay, high bandwidth consumption, and link breakage |
| Hybrid [3] | High computational complexity to construct and maintain the cluster, cluster head, and cluster member. |

The limitations of position-based routing protocols for FANETs are summarized in Table 3.

**Table 3.** Summary of position-based routing protocols and their limitations in FANETs [37].

| Path Strategy | Protocol | Limitations to Adopt in FANETs |
|---|---|---|
| Single-path strategy | Greedy forwarding | • Always seeks progress in transmission distance; thus, it cannot ensure the desired link quality.<br>• Encounters link breakages, blind paths, and routing loops.<br>• Not energy efficient |
| | Compass | • High possibility to trap in routing loops<br>• Not energy efficient |
| | Most forward | • Trapped in local minimum (no further node within transmission range to forward toward the destination)<br>• Encounters higher link breakages and blind paths<br>• Not energy efficient |
| Multipath strategy | Restricted directional flooding | • Deterministic decision to select the direction of broadcasting packets<br>• Broadcast multiple copies of the same packet to the selected direction<br>• Provides excessive overhead and is not energy efficient |
| | Randomized directional flooding | • Randomized decision to select the flooding direction<br>• Provides excessive overhead and high contention<br>• Not energy efficient |
| | Simple Flooding | • Excessive overhead and high contention<br>• Not energy efficient |

According to the above discussion, the features supported by QL-based position-aware routing protocols are listed in Table 4 compared with only position-based forwarding techniques.

**Table 4.** Important features of position-based forwarding techniques and QL-based position-aware routing protocols.

| **Protocol Type** | **Important Features** | | | | | |
|---|---|---|---|---|---|---|
| | **Link Quality Assessment** | **Routing Loops Avoidance** | **Routing Holes Avoidance** | **Energy Holes Avoidance** | **Delay Optimization** | **Blind Path Avoidance** |
| Position-based forwarding techniques | × | × | × | × | × | × |
| QL-based position aware routing protocols | √ | √ | √ | √ | √ | √ |

**Note:** "×": The corresponding feature is not supported; "√": The corresponding feature is supported.

### *1.2. Contribution of This Study*

The key contributions of our study can be summarized as follows:

- We discuss all the realistic mobility models for highly dynamic FANETs based on suitable applications.
- We extensively review recently published QL-based position-aware routing protocols and their advantages and limitations.
- Existing routing protocols are qualitatively compared in terms of their main concepts, key features, performance metrics, and implementation aspects.
- We summarize all the important performance enhancement criteria (in the Lessons Learned section) to design QL-based routing protocols using the position information.
- We identify open issues and research challenges in designing QL-based position-aware routing protocols and their potential research directions in highly dynamic FANETs.

### *1.3. Organization of This Article*

The remainder of this paper is organized as follows: In the next section, we discuss FANETs, their components, their dynamic time-varying topology, and suitable mobility models for FANETs. In Section 3, we briefly review the QL algorithm and its relationship with routing in FANETs. In Section 4, QL-based position-aware routing protocols are extensively reviewed, and their respective advantages and limitations are outlined. In Section 5, existing QL-based position-aware routing protocols are qualitatively compared. In Section 6, the open issues and research challenges associated with the respective potential research directions are discussed. Finally, the paper is concluded in Section 7.

## **2. Flying Ad Hoc Networks**

In this subsection, we briefly discuss FANETs, their components, and the functionalities of each component. We then define the dynamic topology of FANETs. We also discuss the differences of FANETs from other ad hoc networks. Finally, we review the suitable mobility models for FANETs according to the application scenario.

FANETs primarily consist of UAVs that mimic the behavior of swarm intelligence and collaborate with each other and the terminal BS/IoT devices/sensors/GUs or edge–fog–cloud to form an autonomous self-organized multi-UAV communication system. In FANETs, terrestrial devices are swapped with UAVs and can establish communication in any type of emergency without requiring any fixed network infrastructure. In emergency applications, when a terrestrial communication infrastructure or ground sensor network is unavailable, UAVs can be deployed to sense remote areas, utilizing their advantages, such as flexible 3D mobility, fast deployment, and large birds-eye vision. Each UAV in the swarm can sense, execute a computationally intensive task locally or by offloading to the nearest edge server, communicate, cache data, and operate as a router to forward remote

UAV sensing data to the BS for further processing. Thus, a FANET has two major parts: terrestrial and non-terrestrial parts (Figure 1).

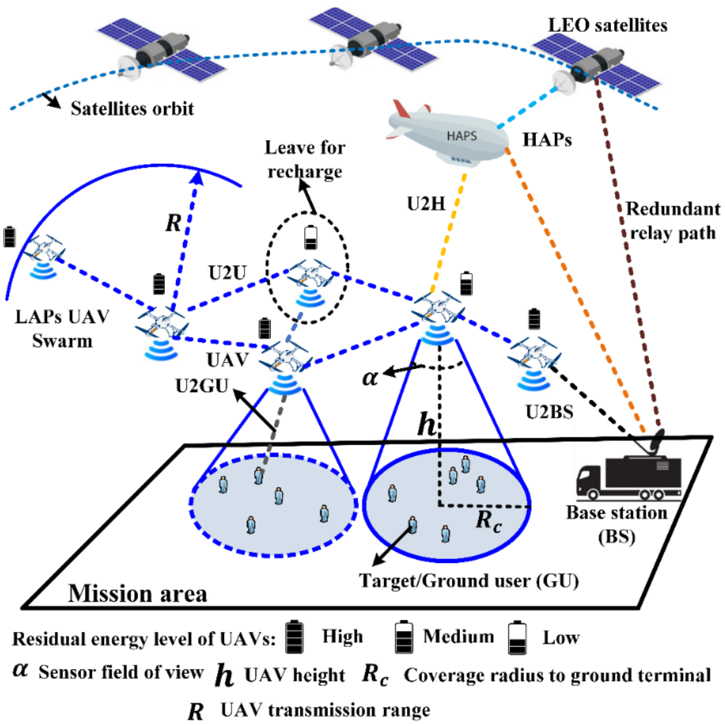

**Figure 1.** FANET and its different components.

The terrestrial section frequently consists of single or multiple fixed or mobile BSs, charging stations, edge-computing servers, GUs, and different types of sensors or IoT devices. The mobile BS can be a ground vehicle with charging stations at the top of the vehicles or edge-computing servers to perform computationally intensive tasks for UAVs. FANETs can also collaborate with existing terrestrial BSs on an on-demand basis to expand the capacity of wireless networks. When the residual energy level of an UAV reaches the minimum threshold level, UAVs can leave the aerial network through an optimal charging-scheduling algorithm to obtain energy replenishment at particular wireless charging stations located at the BS. Instead of wireless charging techniques, UAV batteries can be swapped at the BS before returning to the aerial network. The BS can also function as a gateway to connect FANETs with external wired networks, i.e., the Internet.

The non-terrestrial section consists of a set of homogenous or non-homogenous [1] UAVs operating collaboratively at different altitudes. There are many types of UAVs, and researchers select the UAV type according to their application requirements. In [26,30], the authors classified UAVs according to their size, flying altitude, payload capacity, battery capacity, and endurance time. They classified UAVs into high-altitude platforms (HAPs) and low-altitude platforms (LAPs). HAPs have a high payload capacity, long endurance, and high-energy storage systems. HAPs are quasistatic and are deployed to perform a long-term mission. In this survey, we primarily focus on LAP UAV missions. LAP UAVs can be classified as fixed-wing (small aircraft), rotary-wing (single rotor or multi-rotor), and small tethered balloons. Among them, rotary-wing, particularly multi-rotor UAVs, are mostly used in surveillance and ABS applications as they support vertical take-off and landing, hovering, and provide better formation stability. In addition, multi-rotor UAVs fly in confined areas.

Typically, each UAV has four major modules: flight control, energy management, computation, and communication modules. The flight-control module is responsible for controlling the mobility of UAVs in a 3D space. The mobility information of UAVs in a 3D space can be described by six degrees of freedom (DoFs): surge, heave, sway, pitch, yaw, and roll. The formation controller obtains the optimal mobility control policy by adjusting these six DoFs according to the neighbor UAV mobility (position, velocity, and acceleration) and targets or the GU distribution/mobility within the mission area. It determines the optimal 3D positions and linear/angular velocity for each UAV to achieve both the mission and communication performance. The term surge defines the UAV's forward and backward movement on the X-axis, the sway defines the left and right movement on the Y-axis, and the heave defines up and down movement on the Z-axis. These three terms are used to calculate the 3D Cartesian coordinates of the UAVs, and the rate of change of these 3D coordinates defines their linear velocity. Similarly, the term roll defines side-to-side tilting on the X-axis, the pitch defines forward and backward tilting on the Y-axis, and the yaw defines left and right turns on the Z-axis. These three terms (roll, pitch, and yaw) define the attitude angles and angular velocities. The energy management module produces continuous power for UAVs to generate thrust using UAV rotors and propulsion energy to continue flying in the air. It also supplies energy to perform the computation of different algorithms to execute the mission and communicate with other neighboring UAVs, remote BSs, or GUs with the aid of a communication module. Frequently, the propulsion energy of an UAV is significantly higher than the communication energy. The propulsion power consumption is proportional to the UAV trajectory [41]. Thus, in UAV missions, the joint optimization of UAV trajectory and UAV communication provides better energy efficiency [41]. The computation modules execute algorithms to process the collected data using onboard sensors such as cameras, LiDAR, sound navigation and ranging (SONAR), and inertial measurement units (IMUs). Communication modules consist of wireless radio antennas and wireless interfaces.

There are various communication links in FANETs depending on the mission planning and control methods. These communication links can be classified as air-to-ground (A2G) and air-to-air (A2A) links. Generally, A2G links include UAV to BS (U2BS) and UAV to GU (U2GU). Similarly, the A2A links are UAV to UAV (U2U), UAV to HAPs (U2H), and UAV to low earth orbit (LEO) satellites (U2S). UAVs can directly communicate with satellites [42], particularly with GPS, to localize themselves in global coordinates. LAP UAVs communicate with the BS using U2BS downlinks. U2BS and U2GU links are mostly Wi-Fi links, and they have low costs in terms of delay, transmission power, latency, and path loss in line-of-sight (LoS) scenarios. However, the quality of the A2G links significantly degrades in the no-line-of-sight (NLoS) scenarios, particularly in the urban environment. However, depending on the signal quality and mission design, LAP UAVs can also utilize U2H or U2S uplinks to communicate with the BS as a redundant path [43]. A2A links are mostly considered free-space paths and are dominated by LoS links [1,15]. However, owing to the high mobility and energy limitations of UAVs, these paths encounter high dynamism, Doppler effects, and link breakage.

### 2.1. Dynamic Topology in FANETs

The topology of a FANET can be described as a time-dependent graph, where the connectivity among UAVs is subject to change owing to the mobility of UAVs and their limited battery capacity. Frequently, the FANET mission execution time ($T$) is divided into sufficiently small time slots ($t$) of equal length, and in each time slot, the mobility of the UAVs is considered static. Thus, in FANETs, the network topology can be expressed as a time-dependent undirected graph: $G(t) = (V(t), E(t))$. Here, the vertex $V(t)$ consists of $u_i \in U(t)$ UAVs and single or multiple BSs, and $E(t)$ represents the network edges. In each $t$, if the distance between two UAVs is $d_{ij}(t) \leq R$, where $R$ is the LoS transmission range of each UAV, a direct edge between two UAVs $\left(u_i, u_j\right) \in E(t)$ is considered. In each $t$, UAVs can leave the aerial network to obtain energy replenishment at charging stations located in the BS and rejoin the aerial network after receiving energy replenishment. The sensor coverage radius ($R_c$) of an UAV depends on the UAV height ($h$) and sensor field of view ($\alpha$) (see Figure 1). With increasing $h$, $R_c$ and the probability of obtaining an LoS to the ground terminal increase. Simultaneously, the path loss to the ground terminal increases. Thus, the height of UAVs must be optimized according to the distribution of GUs, mobility of GUs, mission environment, and the application type of FANETs.

### *2.2. Mobility Models in FANETs*

The performance of routing protocols in FANETs is generally evaluated using testbed experiments or simulations in software environments. A testbed enables the performance of the designed routing protocol in real environments to be evaluated. However, owing to the high hardware cost, time requirements, and complexity of constructing large-scale networks with different topologies in testbed scenarios, software simulations are generally preferred for assessing performance. Software simulation-based evaluations of routing protocols in FANETs require a suitable realistic mobility model to define the optimal trajectories of each UAV in the swarm and to specify how their six DoFs (linear and angular velocity, position, attitude angles, acceleration, and direction) change with time in a 3D space.

The mobility models in FANETs should be focused on the type of UAV swarm mission and be suitable for FANET applications such as surveillance, ground target searching and tracking, and search and rescue operations. The mobility model in FANETs should satisfy the following requirements: UAVs should autonomously adjust their flying directions and plan their displacements according to the mobility information of their neighboring UAVs and the location of the target at the ground terminal. Each UAV should maintain a particular safety distance to avoid inter-UAV collisions while simultaneously staying adequately close to ensure QoS in aerial connectivity. Each UAV should be able to join or leave the FANET arbitrarily. To preserve coordination and synchronization in movements, UAVs should continuously adjust their position and velocity according to the mean velocities of neighboring UAVs. The trade-off between maximizing sensing coverage and aerial connectivity [8] and self-healing to the failure of a neighboring UAV should be satisfied. The trajectory of each UAV in a swarm should be smooth, and the moving trajectory of each UAV should maintain fairness in terms of travel distance to maintain a balance in the energy consumption of UAVs within the swarm [44]. Additionally, when UAVs fly over the mission area, they should cover each zone evenly, and repetitive coverage should be avoided as much as possible to improve the search efficiency [44]. Finally, UAVs should be aware of external obstacles when adjusting their mobility. Owing to the above requirements, the mobility models in FANETs are unique and differ from the mobility models proposed for mobile ad hoc networks (MANETs) and vehicular ad hoc networks (VANETs). A realistic mobility model with a proper topology can jointly optimize the mission performance and communication performance.

Several survey papers on FANET routing protocols have comprehensively reviewed mobility models in FANETs. In [30], the authors provided a novel taxonomy for mobility models in FANETs by classifying them into five different categories: random-based, time-based, path-based, group-based, and topology-based. In [31], the authors briefly discussed mobility models for FANETs, including random direction, random waypoint (RWP), reference point group mobility, Gauss–Markov, semi-random circular movement, Paparazzi, and smooth-turn mobility models. However, none of these mobility models adopt the collaborative and cooperative properties of swarm intelligence and they are mostly suggested for MANETs and VANETs [9]. Additionally, the aero dynamic constraints and UAV six DoFs are not taken into account by them [21]. FANETs differ from MANETs and VANETs in terms of their 3D mobility, node density, rate of topological alterations, and energy limitations. Both MANET and VANET nodes have two-dimensional mobility, and VANETs have higher mobility than MANETs. However, in VANETs, nodes are not energy-constrained, and their mobility is limited by roads [45]. Thus, the topology prediction in VANETs is much easier than that in FANETs. In FANETs, UAVs have mobility with six DoFs, the presence of external obstacles, limited energy, system stability to ensure stable links, and restricted trajectories owing to collaborative motion planning. Thus, the mobility models proposed for MANETs or VANETs are not suitable for application to FANETs. Motivated by these limitations of previous studies, we provide a brief review of realistic

FANET mobility models that mimic the characteristics of a swarm according to recently published research articles.

In [9], the authors proposed a novel mobility model for FANETs inspired by boid flocking. They defined the neighbors of UAVs into three categories by dividing the transmission range of UAVs into three zones: repulsive, stable, and attractive. They updated the accelerations, velocities, and positions of UAVs by applying seven different rules: separation, cohesion, alignment, centripetalism, consistency, and synchronization. Each of these rules defines one motion component, and the summation of these rules defines the optimal mobility for each UAV in the swarm. This behavior-based mobility model, inspired by bird flocks, supports both the coverage and connectivity requirements of FANETs with inter-UAV collision avoidance. Thus, it is suitable for performing surveillance, target searching, and rescue operations in emergency scenarios.

Similarly, in [4,10], the authors proposed a virtual force-based mobility model for FANETs by applying four different virtual forces to optimally deploy UAVs as an ABS. The virtual forces they considered to obtain the position and velocity for each UAV in the swarm were attractive forces toward the hot spot areas of GUs, attractive forces toward the isolated GUs, repulsive forces to neighboring UAVs to avoid inter-UAV collisions, and repulsive forces to external obstacles to avoid collision with external obstacles. The sum of these virtual forces defines the acceleration, velocity, flying direction, and position of the UAVs. This virtual force-based mobility model supports optimal coverage of the GUs and simultaneously maintains stable bi-connectivity in the aerial network. Additionally, it can generate a smooth trajectory for UAVs while maintaining fairness in the travel distance between UAVs in the swarm. The adaptive adjustment of the attractive and repulsive force weights can effectively manage swarm stability, connectivity, and obstacle avoidance [11,17].

In [18,19], the authors proposed a mobility model for FANETs inspired by Hooke's law of springs, which is known as the virtual spring-based mobility model. According to Hooke's law, they assumed that the force is proportional to spring deformation. In [18], they proposed a UAV positioning method in a hexagonal pattern to maximize the coverage under QoS in connectivity by defining the attractive and repulsive virtual spring force laws. If the inter-UAV distance crosses the minimum threshold, the spring force is repulsive. In contrast, if the inter-UAV distance crosses the maximum threshold distance, the spring force is attractive. By computing the summation of all virtual spring forces with all neighboring UAVs, each UAV determines the optimal mobility. This type of mobility model is suitable for surveillance missions. Similarly, in [19], the authors utilized a virtual spring force mobility model to maintain a strong aerial mesh network to provide communication in a disaster scenario. Although this type of mobility model can support coverage and QoS in aerial connectivity, it is not well-suited for external obstacle avoidance.

In [12,16,46], the authors proposed a mobility model using an APF. APFs have attractive fields for neighboring UAVs, repulsive fields to maintain the minimum separation distance with neighboring UAVs, and repulsive fields to avoid external obstacles. The fields are computed according to the inter-UAV distance and the distance between the UAV and obstacles. The optimal mobility of UAVs can be obtained by obtaining the negative gradient of the net artificial potential field. This mobility model is suitable for maximizing coverage under connectivity constraints and avoiding external obstacles and inter-UAV collisions. Many algorithms use APFs to maintain the leader–follower topology [47], target searching [46], and predefined trajectory tracking. However, this type of mobility model can easily be trapped in the local minimum problem [12], particularly in highly confined and dense obstacle mission areas. The adaptive adjustment of the attractive and repulsive force constants of APFs can control the optimal node density in FANETs and minimize interference in inter-UAV communication [12].

In [48,49], the authors proposed a distributed pheromone repel mobility model to execute reconnaissance and target-searching missions. Each UAV maintains its pheromone map and scans the mission area according to its map. The UAVs exchange information among themselves to assemble a global pheromone map. The UAVs turn right or left or travel straight according to pheromone smell probabilities. UAVs prioritize regions with a low pheromone smell to explore more in the mission area and reduce repetitive coverage. The limitations of this type of mobility model are that they only consider target exploration; they do not consider the connectivity among UAVs in the aerial network. Additionally, the trajectories of UAVs are not smooth and become complex.

The realistic mobility models discussed above are summarized in Table 5 according to their applications in FANET.

**Table 5.** Summary of realistic mobility models in FANETs according to applications.

| Mobility Model | References | Applications in FANETs |
|---|---|---|
| Boid flocking | [9] | Surveillance, target searching, and communication in an emergency scenario, and simulating the routing and MAC protocol |
| Virtual force | [4,10,11,17] | Surveillance, ABS deployment, ground target searching and tracking, trajectory tracking, and leader–follower topology formation |
| Virtual spring | [18,19] | Aerial base station deployment, surveillance, aerial mesh network deployment, and target searching and tracking. |
| Artificial potential field | [12,16,46,47] | Surveillance, leader–follower topology formation, trajectory tracking, and target searching and tracking |
| Pheromone-based | [48,49] | Ground target searching and tracking, and reconnaissance |

## 3. Q-Learning and Its Relationship with Routing in FANETs

This section provides a brief overview of the effective QL algorithm and its relationship with position-based routing decisions in FANETs.

In QL, agents iteratively adjust their action strategies through the reward that they achieve from the environmental feedback after performing a particular action (Figure 2). The agent selects an action in a particular state according to the reinforcement or simply the previous experience, known as the Q-value. The reinforcement comprises a direct reward and future Q-value expectation. Through reinforcement, agents can evaluate the effectiveness of an action in the current state and perform a better action in the next step. The objective of the agent is to maximize the expectations of the cumulative rewards over the sequential iteration. The advantage of QL is that the reward function can be designed using multiple weighted objectives to achieve multiple goals. Additionally, a suitable strategy for exploration and exploitation can aid in attaining the global optimal solution.

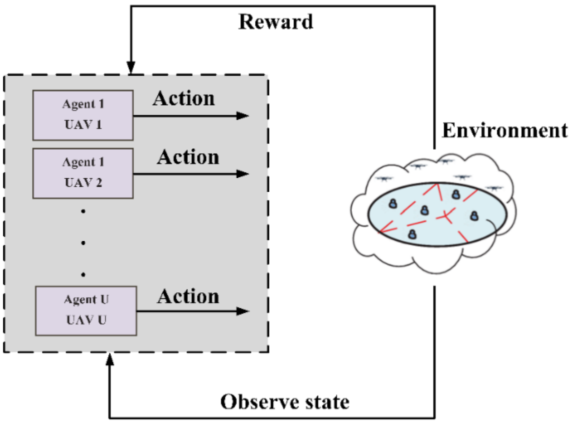

**Figure 2.** State, action, and environment relationship with an agent in the QL model in FANETs.

In FANET routing, the data packets contained by each UAV node (source node) act as agents that function by relaying UAV selections from their neighbor to forward data packets to the desired destination (another UAV or BS). Through this sequential action, the data packets change their states from one UAV to another. This state transition continues until data packets are delivered to the desired destination node. For each action (relay node selection from the neighbor list), the UAV agent (source node) receives a reward or penalty. Gradually, each agent UAV gathers an experience represented by a Q-value that results in an optimal policy in which the cumulative reward is maximized over the iteration. The reward function is designed such that UAVs select an optimal routing path that minimizes the delay and creates a load balance in UAV energy consumption. The penalty mechanism in the reward function is triggered to avoid the local minimum problem if the next state of the data packets does not have any routing path forward toward the destination. It also aids in avoiding the routing loop and minimizes unnecessary detours of data packets during routing.

The QL-based routing decision process can be described as a partial MDP tuple $(S, A, P, R, \alpha, \gamma)$, where a set of states $S \in \{s_{u_i}(t), i = (1, 2, \ldots U)\}$ represents the $u_i \in U$ active UAV node positions at each time $t$, a set of actions $A \in a_{u_{ij}}$ represents the selection of the relay UAV ($u_j$) from the neighbor list of the source UAV ($u_i$) to forward the data packets to the destination, $P$ represents the probability of successful state transition of a particular data packet from one UAV to another UAV, and $R \in r_{u_{ij}}$ is the reward function that can be obtained by a predefined reward function to evaluate the quality of the action (to select the link $u_{ij}$). The Q-values are iteratively updated at each UAV node using the following Bellman equation:

$$Q\left(s_{u_i}, a_{u_{ij}}\right) \leftarrow Q\left(s_{u_i}, a_{u_{ij}}\right) + \alpha\left[r_{u_{ij}} + \gamma \max_{a'_{u_{ij}}} Q\left(s'_{u_{ij}}, a'_{u_{ij}}\right) - Q\left(s_{u_i}, a_{u_{ij}}\right)\right] \tag{1}$$

where the term $\max_{a'_{u_{ij}}} Q\left(s'_{u_{ij}}, a'_{u_{ij}}\right)$ represent the future Q-value expectation in the next state $s'_{u_{ij}}$ after implementing the best action $a'_{u_{ij}}$. $\alpha$ and $\gamma$ represent the learning rate and discount factor, respectively, the values of which are defined within the range $[0\ 1]$. The significance of $\alpha$ is how rapidly the QL algorithm should learn, and $\gamma$ defines how much the QL algorithm learns from its mistake owing to a bad action. The value of $\alpha$ determines the degree to which the newly obtained information overrides the old information, and this parameter controls the convergence of the learning procedure. The value of $\gamma$ controls the importance of future rewards. In FANETs, to obtain a better learning process and more stable Q-value, the values of $\alpha$ and $\gamma$ should be controlled adaptively with respect to the changes in the mobility of UAVs, considering the dynamic topology. The stability of the topology and the degree of change in the mobility of UAVs can be obtained with the aid of a topology control algorithm. Through proper mobility management defined by the topology controller, a UAV swarm can calculate the mobility factor and link duration [22,23]. The parameter link duration defines how long two neighboring UAVs exist within transmission range of each other [21]. The link duration is a function of the UAV transmission range, relative velocity, and distance between two neighboring UAVs [21]. If the mobility factor is high and the link duration is low, the discount factor should be low, and vice versa. Thus, by incorporating a topology controller in the routing decision-making process, UAVs can reduce the number of retransmissions. Similarly, the delay, including the MAC-layer, queueing, and transmission delays, can be obtained from the designed MAC protocol and queuing model [22]. $\alpha$ can be adjusted adaptively according to the delay.

In the QL algorithm, an appropriate balance between exploration and exploitation is necessary to avoid local optima. Exploration refers to the seeking of new actions by the agent (source UAV) to gather new experiences. Exploration can provide good or bad rewards, but it can aid in determining the global optima. In contrast, exploitation involves taking action according to previous knowledge (Q-value). Owing to the limited energy of UAVs, excessive exploration can delay convergence and may include many bad actions that may increase the number of retransmissions. Therefore, proper guidelines are required to explore a better state (relay UAV) from the neighbor list.

## 4. Q-Learning-Based Position-Aware Routing Protocols in FANETs

According to our discussion in Section 1.1, the routing protocols in FANETs are categorized as topology-based and position-based protocols. In this section, we extensively review the QL-based position-aware routing protocols in FANETs, which is a new research trend. The QL-based routing protocols incorporated with position-based forwarding can be considered as a special subcategory of position-based forwarding, where RL is exploited to make the routing decision. As a result, we do not give any further classification for QL-based position-aware routing protocols as they fall into a similar category. The QL-based position-aware routing protocols along with their optimization objective, advantages, and limitations are comprehensively reviewed according to the order of the published year.

### *4.1. Q-Learning-Based Geographic Routing Protocol (QGEO)*

Jung et al. [39] proposed the Q-learning-based geographic routing protocol (QGEO) that outperforms position-based routing protocols in terms of PDR and network overhead in high-mobility scenarios of UAVs. Here, each UAV makes a routing decision in a distributed manner by utilizing the one-hop neighbor node position with the aid of GPS and RL. Each node broadcasts hello packets at a fixed hello-interval that includes mobility information (position, velocity, and direction), current Q-value, link condition (link capacity, interference, and delay), and location error. Unlike the position-based routing protocol, instead of only seeking progress in transmission distance toward the destination, it provides a concept for packet travel speed (PTS) to select the next relay node. The PTS jointly considers the distance progress toward the destination and the channel conditions. The channel condition is computed as the packet travel time (PTT), which consists of the MAC delay, transmission delay, link error, and localization error. The relay node that provides a higher PTS receives more rewards and is selected as the next forwarding node.

PTT ($PTT_{u_{ij}}$) and PTS ($PTS_{u_{ij}}$) between source UAV ($u_i$) and relay UAV ($u_j$) are computed as follows:

$$PTT_{u_{ij}} = \frac{\left(t^{mac}_{u_{ij}} + t^{que}_{u_{ij}} + t^{tx}_{u_{ij}} + t^{p}_{u_{ij}}\right)}{\left(1 - E_{u_{ij}}\right) \times (1 - E_{loc})} \tag{2}$$

and

$$PTS_{u_{ij}} = \frac{d(u_i, BS) - d\left(u_j, BS\right)}{PTT_{u_{ij}}} \tag{3}$$

where $t^{mac}_{u_{ij}}$, $t^{que}_{u_{ij}}$, $t^{tx}_{u_{ij}}$, and $t^{p}_{u_{ij}}$ represent the MAC, queuing, transmission, and propagation delays, respectively, for link $u_{ij}$. $E_{u_{ij}}$ and $E_{loc}$ represent the link and localization errors, respectively, for link $u_{ij}$. $d(u_i, BS)$ and $d\left(u_j, BS\right)$ represent the distance between the source UAV and destination BS and the distance between the selected relay UAV and destination BS, respectively. The parameter $PTS_{u_{ij}} > 0$ means that $u_j$ indicates the progress of distance toward the destination BS, and a higher value of $PTS_{u_{ij}}$ means that the link condition of $u_j$ is good and offers a very small delay.

If a local minimum incident occurs, the corresponding relay node receives a minimum reward as a penalty.

*Advantages*: QGEO overcomes the limitations of position-based routing protocols by considering link conditions and location errors during the next-hop selection process via QL. QGEO can avoid the local minimum by assigning a minimum reward.

*Limitations*: QGEO does not consider the mobility control of the UAV swarm and considers a fixed hello interval (4 Hz) during the simulation. The learning rate and discount factor are not adjusted adaptively with the mobility of the UAV swarm. Additionally, the RE of the UAV node is not considered while making the routing decision. QGEO does not consider the balance between exploration and exploitation strategies to explore better relay UAVs.

### *4.2. Reward Function Learning for QL-Based Geographic Routing Protocol (RFLQGEO)*

Jin et al. [50] proposed the reward function learning for a QL-based geographic routing protocol (RFLQGEO) that provides less retransmission, lower average end-to-end delay, and higher PDR compared with QGEO. RFLQGEO utilizes the inverse RL concept to design the routing decision by exchanging the hello packet, which accelerates the learning process with less communication overhead. RFLQGEO has three components: the location information module, QL routing module, and reward function learning module. The location information module uses hello packets to share the position obtained by the GPS, link condition, residual energy level, link error, location errors, and Q-value with neighboring UAVs. The QL module updates the Q-value using a reward function that jointly considers the distance between two UAVs directed to the destination sink node, PTT including the MAC and transmission delay, link and position error, and void area avoidance feature.

*Advantages*: RFLQGEO outperforms QGEO in terms of designing the reward function for better learning by including the distance progress in the direction of the sink and void area avoidance component in the reward function.

*Limitations*: The limitations of RLQGEO are that it does not consider the adaptive control of the hello interval, mobility control mechanism, and exploration and exploitation method to avoid local optima during the learning process. Additionally, the residual energy of the UAV node is not considered in the reward function.

### *4.3. QL-Based Cross-Layer Routing Protocol (QLCLRP)*

He et al. [51] proposed the QL-based cross-layer routing protocol (QLCLRP), which utilizes the upper confidence bound (UCB) to determine the balance between exploration and exploitation. They proposed carrier-sense multiple access with multi-channel automatic synchronization (CSMA/MAS) to reduce collisions in the MAC layer by maintaining transmission in a round-robin fashion with synchronization. The MAC delay statics are sent to the QL module for better routing decision-making. The Q-value is adaptively updated for better decision-making by updating the learning rate and discount factor according to the mobility of the node and variance of the MAC delay. The reward function includes the difference between the distances between two UAVs projected to the destination and the MAC delay.

*Advantages*: QLCLRP uses the cross-layer concept with an exploration strategy using UCB that enhances the performance of both the MAC and routing layers.

*Limitations*: In QLCLRP, the hello interval frequency is not adaptive. It does not consider mobility control and energy-efficient routing, although the exploration rate was defined by calculating the UCB for each source node considering the number of times a node is selected as a forwarding node and the number of times a source node makes the routing decision.

### *4.4. Multi-Objective QL-Based Routing Protocol (QMR)*

Liu et al. [22] proposed a multi-objective QL-based routing (QMR) protocol for FANETs that jointly optimizes the delay and energy consumption in FANETs while making routing decisions. They proposed a new exploration and exploitation mechanism considering the network condition that includes the PTS, link quality, and relative mobility factor, which aids in avoiding local optima. To make the learning process more efficient and stable, they updated the Q-value by adaptively updating the learning rate and discount factor, utilizing the exponential of the normalized one-hop delay and variation in the neighbor set at two different times, respectively. During exploitation, a better decision is made by the QL module as Q-values are weighted by the link-quality metric. QMR estimates link quality using the expected transmission count (ETX) method [52].

*Advantages*: QMR considers a minimum reward policy when the routing loop, local minimum, and node failure-state occur during relay node selection. It jointly considers delay and energy optimization in the reward function.

*Limitations*: QMR does not consider mobility control and the control of the hello interval. The Q-value is updated without considering the SINR level of the links. Proper mobility control is necessary for QMR, as they provide actual velocity constraints for neighboring UAVs by calculating the PTS to satisfy the deadline PTT during data transmission. A mobility control method can control the relative velocity with neighboring UAVs and maximize the link duration between neighboring UAVs for successful data transmission within PTT.

### *4.5. QL-Based Multi-Objective Fuzzy Routing Protocol (QLMF)*

Yang et al. [53] proposed the QL-based multi-objective fuzzy routing (QLMF) for FANETs, in which the next hop is selected by sending the link and path-related parameters to a fuzzy controller. The link-related parameters are the transmission rate, energy state, and flight status, such as similarity in the flying direction. The path-related parameters are dynamically updated by the RL, which maintains two different Q-values: hop count and successful packet delivery time. Therefore, the fuzzy controller uses three link-related parameters and two path-related parameters as the input membership functions. Subsequently, the parameter is divided into two-level fuzzy linguistic variables that are inferred with the predefined rule base to select a better forwarding node based on the output crisp value after defuzzification.

*Advantages*: QLMF outperforms Q-value-based ad hoc on-demand distance vector routing in terms of hop count and energy consumption.

*Limitations*: Maintaining two different Q-values for each UAV is challenging, and the learning rates are fixed. The delay is calculated considering the number of hops required to reach the destination, which is not appropriate for MAC, queuing, and transmission delays, and must be calculated for precise estimation.

### *4.6. QL-Based Routing Protocol for FANET (Q-FANET)*

Luis et al. [24] proposed an improved QL-based routing protocol for FANETs (Q-FANET), which consists of two sub-modules: QMR [22] and Q-noise+ [53]. In typical QL, the Q-values are updated based on the reward of the most recent episodes. In contrast, considering the mobility in FANETs and dynamic channel conditions, Q-FANET updates the Q-value more precisely by considering the weighted reward for the finite number of last episodes and the SINR level of the selected link. While updating the Q-value, a higher weight is assigned to the recent episodes to obtain a more accurate Q-value. Q-FANET uses the QMR module to select the relay node according to the maximum PTS and to solve the routing hole problem (local minimum). When the routing hole problem occurs, Q-FANET uses the QMR penalty mechanism and allocates the minimum reward to that relay node. Otherwise, a QL module that performs random action (random relay selection) or exploration according to the maximum Q-value is used by adopting the $\in$-greedy policy.

*Advantages*: Q-FANET provides low delay and jitter for a better quality of service in highly dynamic FANETs. It also allocates minimum reward for bad action to overcome the local minimum, routing loop, and node failure-state in FANETs.

*Limitations*: Q-FANET uses the $\in$-greedy policy with random action (relay UAV selection) to balance between exploration and exploitation. However, random actions without proper guidance may provide less reward and produce higher retransmission in FANETs. Additionally, storing the Q-value for the last finite number of episodes requires extra memory consumption for each UAV.

### *4.7. QL-Based Topology-Aware Routing Protocol (QTAR)*

Arafat et al. [23] proposed QL-based topology-aware routing (QTAR) for FANETs, which can make better routing decisions by extending the local view of each UAV using two-hop neighbor information. They updated the Q-values by adaptively adjusting the learning rate based on the exponential normalized two-hop delay and discount factor based on the one-hop current and previous neighbor similarity set of each UAV neighbor. Through this process, QTAR produces a more stable Q-value for better exploration, considering the dynamism in the FANET topology. The reward function includes the delay in terms of the MAC and queuing delay, UAV residual energy level, and PTS for two-hop links. They adaptively adjusted the hello interval based on the minimum link duration among one-hop neighbor UAVs.

*Advantages*: The multi-objective reward function using two-hop neighbor information optimizes the delay and creates a proper load balance in the UAV energy consumption during multi-hop routing. The penalty mechanism in the reward function ensures avoidance of routing loops, routing holes, and local optima.

*Limitations*: QTAR considers velocity constraints that jointly consider distance progress and channel conditions in terms of MAC and queuing delay for two-hop neighbor UAVs to ensure data packet delivery within the deadline PTT. Proper mobility control is necessary to satisfy this PTS constraint; otherwise, the selected relay UAV may leave the transmission range of the corresponding source UAV owing to the change in relative velocity. Maintaining two-hop neighbor information produces a high overhead for each UAV.

### *4.8. Predictive Ad Hoc Routing Protocol Fueled by Reinforcement Learning and Trajectory Knowledge (PARRoT)*

Benjamin et al. [29] proposed predictive ad hoc routing fueled by reinforcement learning and trajectory knowledge (PARRoT) for FANETs to perform the collaborative mission of a UAV swarm. In PARRoT, each UAV exchanges hello packets with a one-hop neighbor, which includes a unique UAV address, hello packet sequence number, current position, predicted future position, reward, and cohesion value. Based on the hello packet transmission and reception in reverse order (from the destination to the source node), each node updates the link-quality status represented as a Q-value. To address the dynamic topology in FANETs, the discount factor is calculated by considering two dynamic parameters: link duration and cohesion value. The link duration is calculated using a mobility controller that utilizes the trajectory knowledge (relative velocity, relative position, and transmission range) of two neighboring UAVs. The cohesion value is calculated based on the variation in the neighbor set at two different times. The neighbor link that provides a high link duration and cohesion value receives more rewards. To make the routing decision, each source node selects a neighbor link based on the maximum Q-value.

To simulate the routing behavior, they considered both generic and realistic mobility models. For instance, they utilized the distributed dispersion detection (DDD) mobility model [54], which is suitable for exploring plume sources in disaster scenarios. In addition, they considered the dynamic cluster hovering (DCH) method [55], in which a UAV swarm dynamically adjusts its locations to provide network coverage of ground-based vehicles.

*Advantages:* PARRoT introduces relative trajectory knowledge between two neighboring UAVs to make routing decisions in FANETs. Owing to the knowledge of relative mobility information, it selects a more stable neighbor link in FANETs and provides the highest survival time of the link to complete data transmission.

*Limitations:* The residual energy of the neighbor node and routing loop are not considered in routing decision making. When making the routing decision, PARRoT always exploits the link with the maximum Q-value; thus, no exploration criteria are considered.

### *4.9. Simulated Annealing Inspired Q-Learning (SAIQL)*

Sugranes et al. [56] proposed simulated annealing-inspired Q-learning (SAIQL) for FANETs that utilizes an adaptive learning rate operated through heuristic simulated annealing optimization (SAO), where the temperature parameter ($T$) captures the influence of the UAV's mobility to update the Q-value. SAIQL considers a piecewise linear mobility model, where UAVs maintain piecewise linear motion over a time slot whose duration is exponentially distributed. In each time slot, the UAVs maintain constant velocity and direction, which may vary in the next time slot. The learning rate is obtained using $T$ of the

SAO, which promptly adapts to the UAV's average velocities during each time slot. SAIQL reduces the energy consumption of UAVs and increases the PDR of FANETs.

*Advantages*: SAO at high temperatures has more aggressive exploration rates, and the algorithm learns gradually toward the optimal decisions over time by cooling the temperature that minimizes the end-to-end transmission energy of UAVs. SAIQL avoids routing loops as intermediate UAVs cannot be included in the end-to-end routing path more than once, and each UAV maintains an updated Q-table with information regarding the visited UAV nodes.

*Limitations*: During exploration, a new action is selected randomly, which may produce unnecessary detours and delay the convergence of the algorithm. A specific guide is necessary to select a new action during exploration, such as a set of UAV nodes with sufficient residual energy, which indicates the distance progress toward the destination and satisfies the PTT to deliver the data successfully within the deadline.

## 5. Comparison of QL-Based Position-Aware Routing Protocols

In this section, the QL-based position-aware routing protocols discussed in Section 4 are discussed in terms of their objectives and innovative features. Additionally, they are qualitatively compared.

The objective and innovative features of the QL-based position-aware routing protocols are summarized in Table 6.

**Table 6.** Objective and innovative features of QL-based position-aware routing protocols.

| Protocol | Objective | Innovative Features |
|---|---|---|
| QGEO [39] | Utilizes MAC delay, link error, and localization error in QL to determine the link quality and avoid the local minimum problem. | Introduces the concept of PTS and PTT in QL-based position-aware single path forwarding. |
| RFLQGEO [50] | Utilizes MAC delay, link error, localization error, and distance progress toward destination to determine the forwarding node. | Introduces inverse reinforcement learning by exchanging hello packets to estimate the link Q-value as a link quality metric. |
| QLCLRP [51] | Design a cross-layer adaptive QL-based routing protocol to minimize delay, link error, and localization error. | Introduces exploration and exploitation technique in QL model using UCB. Moreover, adaptively controls the learning rate and discount factor according to the node mobility and delay. |
| QMR [22] | Jointly minimizes the delay and energy by designing a multi-objective reward function in adaptive QL. | Introduces the exploration and exploitation strategy in QL-based routing decisions according to the degree of mobility with neighbor UAVs. |
| QLMF [57] | Utilizes QL and fuzzy logic to select the optimal path in FANETs by jointly considering two path-related parameters (hop count, packet delivery ratio) and three link parameters (transmission rate, energy state, and flight direction). | Designs an integrated framework using fuzzy logic and QL to select the optimal forwarding path. |
| Q-FANET [24] | Designs an adaptive QL model to minimize jitter, end-to-end delay, and latency in highly dynamic conditions in FANETs with local minimum avoidance. | Q-values are updated more precisely by considering the weighted reward for the finite number of last episodes and the SINR level of the link. |
| QTAR [23] | Designs a multi-objective reward function to optimize delay, energy consumption, to avoid routing loop, and blind path challenges in FANETs. | Introduces PTS and PTT up to two-hop neighbor links to extend the local view of each UAV for early convergence. |
| PARRoT [29] | Designs an adaptive QL model by exchanging hello packets with neighboring nodes to select more stable paths in FANETs. | Introduces relative trajectory knowledge (link duration, and cohesion metric) to adaptively train the QL model. |
| SAIQL [56] | Designs an adaptive QL model to select the energy and delay optimal relay node in FANETs while avoiding routing loop. | Introduces SAO to control the balance between exploration and exploitation rate in FANETs. |

The mobility model, localization technique, neighbor information, MAC protocol type, exploration strategy, and simulation tool are presented in Table 7. Most of the routing protocols consider generic mobility models such as Gaussian Markov and RWP mobility models for simulating the designed routing protocols. Only PARRoT considers realistic mobility models for UAV swarm missions by adjusting UAV locations according to the mobility of the ground vehicle to provide better network coverage. All of them utilize GPS to detect locations in the global frame. All routing protocols except QTAR utilize one-hop neighbor information to select relay nodes. By utilizing two-hop neighbor information, QTAR extends the local view of each agent and avoids link breakage and blind-path challenges in FANETs. Most routing protocols consider the IEEE 802.11g/n default MAC standard to define the physical layer standard. To achieve a balance between exploration and exploitation in the QL model, some algorithms consider the exploration strategy based on UCB, $\in$-greedy strategy, one-hop PTS, and two-hop PTS and integrate the metaheuristic algorithm (SAO) with the QL model. Network Simulator 3 (NS-3), MATLAB, and WSNet are popular tools used to simulate QL-based position-aware routing protocols.

The path selection strategy, local minimum avoidance, adaptability in the hello interval, link stability, localization error consideration, and scalability features are summarized in Table 7 (in the continued table). All routing protocols utilize a single-path strategy to discover the multihop routing path to reduce the number of broadcasts in energy- and bandwidth-constrained FANETs. They avoid the local minimum problem in the position-based forwarding technique by assigning the minimum reward to the corresponding relay nodes that have no further nodes to forward the data packets toward the destination. The SAIQL avoids the routing loop problem by saving the visited UAVs in the end-to-end path, and none of the forwarding UAVs select the relay UAV that has been selected previously in the end-to-end path. QTAR optimizes the hello interval in FANETs by adaptively selecting its value according to the minimum link duration found within a one-hop neighbor. This technique adaptively sets the hello interval and controls the control overhead. We consider the stability of the selected links to be higher in QMR, QTAR, and PARRoT. This is because both QMR and QTAR consider the PTS and residual energy levels in the reward function when selecting the relay node. PARRoT introduces a mobility prediction mechanism by calculating the link duration and cohesion metric according to the relative trajectory knowledge of neighboring UAVs. Thus, it selects a more stable path by considering high-mobility FANETs. Only QGEO, RFLQGEO, and QLCLRP consider the node localization error when selecting the relay UAV. We consider a low scalability for QLMF because it updates two different Q-values in the learning process and send them to the fuzzy logic controller for decision making. We consider high scalability for QGEO, RFLQGEO, QMR, and PARRoT because they maintain only single Q-value to update the quality of neighbor links. We consider medium scalability for QTAR because it utilizes two-hop neighbor information to make a better optimal decision; however, keeping two-hop information at each UAV produces higher overhead. Similarly, we consider medium scalability for SAIQL because it utilizes the SAO-based metaheuristic algorithm to update the Q-values for each neighbor link, which is an iterative process.

The end-to-end delay optimization, UAV energy efficiency, control overhead, PDR, and adaptive learning features are summarized in Table 7 (in the continued table). We observed that QGEO, RFLQGEO, and QLCLRP consider only the MAC delay to select relay node links. Additionally, they consider the progress of the transmission distance toward the destination node to select the relay link. Among them, QMR, Q-FANET, and QTAR jointly considers the MAC, queuing, and transmission delays to select the relay node. QMR, QTAR, SAIQL, and QLMF consider the UAV residual energy level to select the relay node as part of their objective in the reward function. Thus, QMR, QTAR, SAIQL, and QLMF produce balance in energy consumption, which prolongs the lifetime of FANETs. QTAR provides a higher control overhead than others because it uses two-hop neighbor information. We can consider medium control overhead size for QLMF because it requires storing two different Q-values. Similarly, we consider medium control overhead for PARRoT and

SAIQL because they learn the optimal policy by sharing data packets with their neighbors. QMR and QTAR give higher PDR compared to others because they choose more stable links by considering PTS and link residual energy. Similarly, PARRoT also gives higher PDR in highly dynamic FANETs because it updates the discount factor and Q-values of the neighbor link according to the relative trajectory knowledge (by calculating the predictive link duration). To adaptively train the QL model according to the mobility dynamism in FANETs, as discussed in Section 4, several algorithms adaptively set the learning rate and discount factor. QGEO considers only two different discount factor values according to inter-UAV distances. In contrast, QMR controls both the learning rate and discount factor according to the normalized one-hop delay and changes in the neighbor sets. The QTAR algorithm adopts the same technique but considers two-hop delay to control the learning rate. As a result, it can produce a more precise Q-value and adopts the dynamical topology behavior more adaptively compared to others.

**Table 7.** Qualitative comparison of QL-based position-aware routing protocols.

| Protocol | Year | Mobility Model | Localization | Neighbor Information | MAC Protocol | Exploration Strategy | Simulation Tool |
|---|---|---|---|---|---|---|---|
| QGEO [39] | 2017 | Gaussian Markov | GPS | One-hop | IEEE 802.11g | × | NS-3 |
| RFLQGEO [50] | 2019 | Gaussian Markov | GPS | One-hop | IEEE 802.11g | × | NS-3 |
| QLCLRP [51] | 2019 | - | - | One-hop | CSMA/MAS | UCB | WSNet |
| QMR [22] | 2020 | - | GPS | One-hop | IEEE 802.11 DCF | One-hop PTS | WSNet |
| QLMF [57] | 2020 | Random | - | One-hop | - | × | MATLAB |
| Q-FANET [24] | 2021 | RWP | GPS | One-hop | IEEE 802.11 DCF | ∈-greedy | WSNet |
| QTAR [23] | 2021 | Gaussian Markov | GPS | Two-hop | IEEE 802.11n | Two-hop PTS | MATLAB |
| PARRoT [29] | 2021 | RWP, DDD, and DCH | GPS | One-hop | IEEE 802.11g | × | OMNeT++ |
| SAIQL [56] | 2021 | Piecewise linear mobility | - | One-hop | - | SAO | - |

| Protocol | Path Strategy | Local Minimum Avoidance | Adaptive Hello Interval | Link Stability | Localization Error Consideration | Scalability |
|---|---|---|---|---|---|---|
| QGEO [39] | Single | Yes | No | Medium | Yes | High |
| RFLQGEO [50] | Single | Yes | No | Medium | Yes | High |
| QLCLRP [51] | Single | Yes | No | Medium | Yes | Medium |
| QMR [22] | Single | Yes | No | High | No | High |
| QLMF [57] | Single | - | No | Medium | No | Low |
| Q-FANET [24] | Single | Yes | No | Medium | No | Medium |
| QTAR [23] | Single | Yes | Yes (Minimum LD) | High | No | Medium |
| PARRoT [29] | Single | - | No | High | No | High |
| SAIQL [56] | Single | Yes | No | Medium | No | Medium |

| Protocol | End-to-End Delay Optimization | UAV Energy Efficiency | Control Overhead | PDR | Adaptive Learning |
|---|---|---|---|---|---|
| QGEO [39] | Yes (MAC delay) | No | Low | Medium | Partial |
| RFLQGEO [50] | Yes (MAC delay) | No | Low | Medium | No |
| QLCLRP [51] | Yes (MAC delay) | No | Low | Medium | Yes |
| QMR [22] | Yes (MAC and queuing delay) | Yes | Low | High | Yes |
| QLMF [57] | Yes | Yes | Medium | Medium | No |
| Q-FANET [24] | Yes (MAC and queuing delay) | No | Low | Medium | No |
| QTAR [23] | Yes (MAC, queuing, and transmission delay) | Yes | High | High | Yes |
| PARRoT [29] | × | No | Medium | High | Yes |
| SAIQL [56] | × | Yes | Medium | Medium | Yes |

**Note:** "-": Information is not provided in the corresponding article; "×": The corresponding feature is not supported; DCF: distributed coordination function; LD: link duration.

## 6. Lessons Learned, Open Issues, and Research Challenges

In this section, according to our discussion in Sections 4 and 5, we provide important lessons learned along with the possible performance enhancement techniques of the reviewed protocols. Then, we discuss important open issues and research challenges with the potential research directions for designing QL-based potion-aware routing protocols in FANETs.

### *6.1. Lessons Learned*

In this subsection, the lessons learned are addressed in brief.

#### 6.1.1. Precise Link SINR Calculation

To make a routing decision, SINR needs to be calculated precisely. In FANETs, the SINR depends on the effect of channel medium, transmission power, inter-UAV distance, node density, mission environment, and UAV mobility. The precise calculation of SINR can boost up the performance of the MAC and routing layers. Thus, to improve the SINR performance, joint mobility control and resource allocation need to be allocated optimally [12,35]. Better SINR ensures high link quality and low packet error rate, which improves the routing performance significantly [23]. Thus, cross-layer design is required in FANETs.

#### 6.1.2. Precise Link Delay Calculation

Link delay for one-hop transmission depends on the MAC delay to access the channel, the queuing delay due to the limited buffer size of each UAV [58], the propagation delay due to inter-UAV distance, and the processing delay in the UAV processor. Additionally, node density is also an important factor that can affect the delay because high node density increases the competition to access the shared medium and very low node density increases the inter-UAV distance. Thus, routing protocols should jointly consider all the parameters to improve the network delay. In [20], the authors adaptively control the node congestion window according to node density in FANETs to precisely calculate the delay. Furthermore, in [22], the authors use an exponentially weighted moving average to precisely calculate the link delay for each UAV.

#### 6.1.3. Q-Learning Algorithm Convergence

The main concern of QL-based optimization is algorithm convergence and lack of training samples in a real-world environment. FANETs have limited energy and communication resources; thus, the algorithm should have faster convergence to optimal decision making. Additionally, it should adaptively learn the optimal policy by dealing with dynamic topology in FANETs. To overcome this issue, QMR [22] and QTAR [23] utilizes the PTS metric to choose the relay UAV. PTS is the ratio between the distance progress toward the destination and the link delay offered by the corresponding relay UAV. Performing the exploration in the initial stage of learning by using the PTS metric helps to overcome the lack of training samples, and it accelerates the algorithm's convergence to optimal decision making. However, exploration based on the PTS may not be effective in highly dynamic FANETs because it predicts the mobility of UAVs only based on the relative distance. In FANETs, the UAV mobility depends on relative distance, relative velocity, and flying direction (six DoFs), which can be precisely calculated only using the predictive link duration metric [21,29,59]. Thus, performing the exploration or exploitation based on the link duration metric should give better link stability and routing performance in FANETs. In addition, a proper strategy is required to perform the exploration and exploitation to obtain better reward.

#### 6.1.4. Reward Function Design

Because the reward function reinforces the algorithm convergence, designing a good reward function is very important to improve the routing performance. The QMR [22] and QTAR [23] jointly consider the link delay and residual energy level of UAVs in their multi-objective reward function and achieve significant performance improvement for PDR, end-to-end delay, and balance in energy consumption. However, designing the reward function considering path stability, delay, and UAV residual energy may provide better routing performance in FANETs.

#### 6.1.5. Self-Healing and Robustness

While performing the mission in a complex dynamic environment, UAVs might lose connection with their neighbors and could be disconnected from the swarm owing to limited energy, hardware failure, inter-UAV collision, or collision with other obstacles. This problem can be solved effectively by designing a self-healing flocking algorithm [10]. The flocking control algorithm generates the optimal mobility and formation for each UAV in a swarm by interacting with its local neighbors, GUs, and dynamic obstacles with the help of their onboard sensors [20]. Thus, joint mobility control and routing can be an interesting research idea to improve both communication performance and mission performance.

#### 6.1.6. Avoiding Routing Holes, Loops, and Energy Holes

Routing holes (local minimum), routing loops, and energy holes are the common problems in position-based forwarding in high-mobility FANETs. The routing holes can be avoided by allocating minimum reward to the relay UAVs owing to taking the bad action by designing an intelligent reward function [22–24,39]. The consideration of residual energy level in the reward function produces balance in energy consumption, which avoids the energy holes. SAIQL [56] tracks the relay UAVs in the end-to-end path so that none of the UAVs are selected more than once to avoid a routing loop in FANETs.

#### 6.1.7. Topology Prediction Accuracy and Localization Error

In position-based forwarding, location error is an important issue which decreases the routing protocol performance significantly. In FANETs, UAVs usually exchange hello packets with one-hop or two-hop neighbors to update their mobility (position and velocity) and neighbor list [21,22]. The hello interval defines how frequently UAVs broadcast their location information with their neighbors. A low hello interval gives an up-to-date position, and it also produces higher control overhead. In contrast, the high hello interval reduces the control overhead, but it reduces the location accuracy. Similarly, delay in hello broadcasting also causes UAVs to record the inaccurate position of their neighboring UAVs. Thus, the hello interval should be controlled according to the degree of mobility changes in FANETs, which can be estimated by predictive link duration [23]. This is because predictive link duration is the function of relative velocity, relative distance, transmission power, and flying directions in 3D space [29].

### *6.2. Open Issues and Research Challenges*

In this subsection, according to the above discussion, we address important open issues and research challenges to design more effective routing protocols for FANETs.

#### 6.2.1. Adopting the Appropriate Channel Model

To simulate the routing protocols in FANETs, it is necessary to adopt an appropriate channel model for all links such as U2U, U2BS, and U2GU links. With the suitable channel model considering the mission environment and type of wireless links, it is possible to estimate accurate SINR, data rate, delay, and packet error rate in FANETs [1,23]. These parameters also help to make better adjustments to mobility and transmission power and improve the performance of MAC and routing protocols. Usually, owing to the advantages of 3D positioning adjustment in high altitude, the U2U links are treated as free-space paths and U2BS links are dominated by the LoS. However, U2U links face dynamism owing to the high mobility of UAVs, and path loss mostly depends on the inter-UAV distance, weather condition, beam pointing error, and the density of nodes. Depending on the mission environment and distribution of GUs, the U2GU links create the probability of having both NLoS and LoS cases, which entirely depends on downlink antenna elevation, azimuth angle, and transmission power toward the GU [4,60]. Here, all the routing protocols only consider the U2U links to the channel model. In [61], the authors discuss the millimeter-wave-based A2A and A2G channel models for UAV swarms, and consider different antenna types that are specially designed by taking into account the UAV platform size, power, and payload constraint. A comprehensive study on the A2G channel model for UAV-based communication is given in [62].

#### 6.2.2. Realistic UAV Mobility Model

According to our survey, all routing protocols except PARRoT [29] consider generic mobility models, such as RWP and Gaussian Markov mobility models. However, according to the discussion in Section 2.2, the mobility models in FANETs should be application dependent and should adopt the behavior of swarm intelligence to achieve realistic results in the simulation environment. Mobility control algorithms, such as boid flocking [9], virtual force [10,11,17], virtual spring [18,19], and APF [12,16,46] produce a realistic mobility model for a UAV swarm in a software simulation environment considering the type of mission. Thus, designing and evaluating routing protocols that consider a realistic mobility model can be an interesting research concept.

#### 6.2.3. Localization to Predict Dynamic Topology

The accuracy of the localization techniques accelerates autonomous execution and enhances the robustness of FANETs. An accurate localization technique supports many core network services such as topology control, collision avoidance, trajectory planning, and routing. In QL-based position-aware routing protocols, location accuracy is a key factor for predicting a better topology. GPS is the most common localization system used in FANETs owing to its wide coverage, flexibility, and convenience. In wide and open outdoor environments, GPS signals are adequately received, and UAVs can localize them in global coordinates. However, in many circumstances, such as bad weather and long-distance radio communication, the GPS signal may be absent (indoor) or inadequate owing to several causes, such as urban NLoS scenarios and enemy-controlled airspaces.

In such scenarios, infrastructure-independent cooperative localization (CL) can be used, where a group of mobile UAVs with communication capabilities use relative distance measurement via different types of range-based and range-free methods to jointly estimate the position and orientation of all UAVs. Optimization techniques such as semidefinite programming [63], gray wolf optimization (GWO) [3], particle swarm optimization (PSO) [13], and multi-dimensional scaling [64] aid in estimating the accurate location with less measurement in a noisy scenario. The single-shot CL incorporated with filtering techniques, such as extended Kalman filtering [65,66] and Monte Carlo localization [5] can produce continuous state estimation utilizing an onboard sensor, such as an odometer and IMU, which provides better localization accuracy. Ultrawideband localization techniques provide better noise-free ranging in an NLoS environment [5]. Recently, the prediction of location and mobility were well studied in FANETs to improve the routing performance in highly mobile FANETs [67]. In addition, according to our earlier discussion, an UAV formation controller updates the mobility of UAVs in the next state based on the mobility in the current state. Thus, QL-based MDP formulation can be used to design a routing protocol with mobility prediction.

#### 6.2.4. Trade-Off between Exploration and Exploitation

Exploration is an attempt to discover a new state in the search space that may provide a better reward compared with the existing experience of an RL agent. Exploitation refers to performing the best action according to existing experience. Exploration aids in determining the global optimal solution. However, during exploration, the action performed might be good or bad because excessive exploration may produce unnecessary detours, retransmissions in FANETs, and delay the convergence of the algorithm. Therefore, in FANET routing decision making using RL, a strategy is required to balance the trade-off between exploration and exploitation to attain the global optima.

Some RL algorithms consider $\in$-greedy [24] and UCB [51] strategies to control the exploration rate. However, in the $\in$-greedy strategy, the exploration rate depends on the parameter $\in$, which is frequently approximately 10%, resulting in a very low exploration rate. The UCB strategy can control the exploration rate by jointly considering the sum of the average cumulative reward and number of times a specific action is selected within a specific time. In [22], the authors reported that the exploration rate should be controlled according to the network condition and degree of mobility changes in FANETs, instead of exploration based on time. This is logical because when the relative neighbor state is stable, UAVs can exploit according to the existing Q-value. Otherwise, when the relative neighbor state is not stable, UAVs can perform exploration according to the predicted link duration with the neighbor links to achieve a more stable routing path.

Metaheuristic algorithms, such as GWO and SAO, are very efficient at exploring the search space to determine the global optimal solution [68]. The hybrid mechanism with metaheuristics and RL can provide a better solution for FANET routing decision making, as RL agents can decide when to explore or exploit. In [18], the authors used SAO to control the exploration rate defined by the temperature parameters according to the mobility in the FANETs.

### 6.2.5. Control Overhead Minimization

Increasing the hello packet broadcast frequency can aid in discovering the updated location of neighboring UAVs in FANETs. However, a low hello interval produces a high control overhead in FANETs, which may consume a significant amount of bandwidth and energy for UAVs. UAV swarms maintain a stable topology by utilizing a topology-control algorithm. However, in a few scenarios such as obstacle avoidance or formation splitting owing to the mission demand, the neighbor relationship changes in the UAV swarm. Therefore, the swarm can sense the degree of mobility and the degree of neighbor intimacy as the relative distance or link duration, according to which it should control the hello interval adaptively. A specific guideline for controlling the adaptive hello interval according to changes in the neighbor relationship is provided in [16,21].

### 6.2.6. Cross-Layer Design

In FANETs, the link delay, SINR level, link reliability, UAV residual energy, and relative mobility are the key factors in defining link stability. The optimal transmission power allocation in the physical layer and optimal physical resource allocation, such as the frequency or time slots in the MAC layer, control the SINR and throughput of the links. Joint consideration of these constrained resources can significantly improve the performance of the routing layer (relay selection) because they are highly coupled. Designing a cross-layer routing protocol with optimal resource allocation, such as transmission power according to the changes inter-UAV distances indicated by a topology controller [12], MAC layer frequency, or time slots, can be an interesting research direction.

### 6.2.7. Precise Calculation of Energy Consumption

In FANETs, the energy consumption cost of UAVs depends on the power consumption for propulsion and communication to transmit and receive data from neighboring UAVs and GUs [41]. However, the propulsion power of UAVs consumes significantly more energy than the communication energy cost [9]. All of our reviewed routing protocols only consider the communication energy when calculating the energy cost. For a realistic performance, the energy cost should be obtained by considering both propulsion and communication power. Appropriate energy consumption defines the presence of UAVs in the aerial network and defines the accurate node density, which is directly related to communication performance. The propulsion power is proportional to the UAV trajectory.

Thus, during a collaborative mission, the trajectory should be optimized and smooth, and all UAVs should travel approximately the same distance to execute the mission [44]. Additionally, the propulsion energy cost depends on the type of UAV deployed to execute the mission. A recent survey discussed the propulsion energy model according to the type of UAV [69].

#### 6.2.8. Routing for Space–Air–Ground Integrated Network (SAGIN)

Recently, fifth- and sixth-generation (6G) technologies have gained interest in both academia and industry owing to their large spectrum resources, high data rates, and long-range transmission requirements. The space–air–ground integrated network (SAGIN) is introduced in 6G technologies to deliver global coverage and ubiquitous services for remote end-users in depopulated areas (desert and ocean), ships, LAP UAV swarms, and HAPs, which cannot be served by general ground BSs for geographical reasons. Owing to the periodic high mobility of LEO satellites, the mobility of LAP UAV swarms and quasi-stationary HAPs in determining the optimal routing in SAGIN becomes very complex. The concept of a store carry and forward [32] routing mechanism is required because the routing path is entirely time dependent owing to the periodic movement of LEO satellites. Here, an intelligent routing protocol is required that can route data to the remote control center by determining an optimal path in terms of delay and energy cost under the threshold cache memory of the LAP UAV swarm and HAPs. In addition, an accurate channel model and a delay model are required to define the link SINR, link delay, and data transmission rate [42,70]. This area is still new, and to achieve the above objective, more focus is required to design the SAGIN network in the simulation environment. A resourceful guideline for implementing this concept is provided in [42,70].

## 7. Conclusions

In this article, we reviewed the existing QL-based position-aware routing protocols for FANETs along with their advantages and limitations according to the characteristics of the dynamic topology. We found that incorporating QL with position-based routing protocols significantly improves the routing performance in terms of energy consumption, end-to-end delay, local minimum avoidance, and routing loop avoidance. Additionally, the QL technique improves the PDR, minimizes the control overhead, and provides tolerance to localization error. The surveyed protocols were qualitatively compared in terms of their objectives, innovation features, and several important performance metrics. In addition, we discussed important performance improvement criteria such as precise SINR, delay calculation, multi-objective reward function, self-healing, and robustness. From our comparative discussion, it was inferred that researchers or engineers can make a choice of an appropriate routing protocol by taking not only their target applications but also their primary performance metrics. Our comparative results and relevant discussion will help them to make such a choice more effectively. In addition, we discussed several important open issues and research challenges along with their potential research directions for improving the routing performance in FANETs. We also found that making the routing decision jointly consider the path stability, PTS, and residual energy level of UAVs enhances the routing protocol performance in highly dynamic FANETs.

**Author Contributions:** The individual contributions of the authors are as follows. M.M.A. reviewed, analyzed, and summarized the Q-learning-based routing protocols for FANETs. S.M. directed the research and contributed to the review and analysis of the routing protocols. The paper was drafted by M.M.A. and subsequently revised and approved by S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by a research fund from Chosun University (2021).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors wish to thank the editor and anonymous referees for their helpful comments that improved the quality of this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. You, W.; Dong, C.; Cheng, X.; Zhu, X.; Wu, Q.; Chen, G. Joint Optimization of Area Coverage and Mobile-Edge Computing with Clustering for FANETs. *IEEE Internet Things J.* **2021**, *8*, 695–707. [CrossRef]
2. Huda, S.M.A.; Moh, S. Survey on computation offloading in UAV-Enabled mobile edge computing. *J. Netw. Comput. Appl.* **2022**, *201*, 103341. [CrossRef]
3. Arafat, M.Y.; Moh, S. Bio-inspired approaches for energy-efficient localization and clustering in uav networks for monitoring wildfires in remote areas. *IEEE Access* **2021**, *9*, 18649–18669. [CrossRef]
4. Wang, H.; Zhao, H.; Wu, W.; Xiong, J.; Ma, D.; Wei, J. Deployment Algorithms of Flying Base Stations: 5G and Beyond With UAVs. *IEEE Internet Things J.* **2019**, *6*, 10009–10027. [CrossRef]
5. Guler, S.; Abdelkader, M.; Shamma, J.S. Peer-to-Peer Relative Localization of Aerial Robots With Ultrawideband Sensors. *IEEE Trans. Control Syst. Technol.* **2020**, *29*, 1981–1996. [CrossRef]
6. Kumar, A.; Sharma, K.; Singh, H.; Naugriya, S.G.; Gill, S.S.; Buyya, R. A drone-based networked system and methods for combating coronavirus disease (COVID-19) pandemic. *Futur. Gener. Comput. Syst.* **2021**, *115*, 1–19. [CrossRef] [PubMed]
7. Yang, H.; Ye, Y.; Chu, X.; Sun, S. Energy Efficiency Maximization for UAV-Enabled Hybrid Backscatter-Harvest-then-Transmit Communications. *IEEE Trans. Wirel. Commun.* **2021**, 1. [CrossRef]
8. Cheng, X.; Dong, C.; Dai, H.; Chen, G. MOOC: A Mobility Control Based Clustering Scheme for Area Coverage in FANETs. In Proceedings of the 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks"(WoWMoM), Chania, Greece, 12–15 June 2018. [CrossRef]
9. Wang, B.; Sun, Y.; Do-Duy, T.; Garcia-Palacios, E.; Duong, T.Q. Adaptive d-Hop Connected Dominating Set in Highly Dynamic Flying Ad-hoc Networks. *IEEE Trans. Netw. Sci. Eng.* **2021**, *4697*, 2651–2664. [CrossRef]
10. Zhao, H.; Wang, H.; Wu, W.; Wei, J. Deployment algorithms for UAV airborne networks toward on-demand coverage. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 2015–2031. [CrossRef]
11. Zhao, H.; Liu, H.; Leung, Y.W.; Chu, X. Self-Adaptive Collective Motion of Swarm Robots. *IEEE Trans. Autom. Sci. Eng.* **2018**, *15*, 1533–1545. [CrossRef]
12. Xu, W.; Xiang, L.; Zhang, T.; Pan, M.; Han, Z. Cooperative Control of Physical Collision and Transmission Power for UAV Swarm: A Dual-Fields Enabled Approach. *IEEE Internet Things J.* **2021**, *4662*, 1–15. [CrossRef]
13. Arafat, M.Y.; Moh, S. Localization and Clustering Based on Swarm Intelligence in UAV Networks for Emergency Communications. *IEEE Internet Things J.* **2019**, *6*, 8958–8976. [CrossRef]
14. Ruan, L.; Li, G.; Dai, W.; Tian, S.; Fan, G.; Wang, J.; Dai, X. Cooperative Relative Localization for UAV Swarm in GNSS-Denied Environment: A Coalition. *IEEE Internet Things J.* **2021**, *XX*. [CrossRef]
15. Xing, N.; Zong, Q.; Dou, L.; Tian, B.; Wang, Q. A Game Theoretic Approach for Mobility Prediction Clustering in Unmanned Aerial Vehicle Networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 9963–9973. [CrossRef]
16. Dai, F.; Chen, M.; Wei, X.; Wang, H. Swarm Intelligence-Inspired Autonomous Flocking Control in UAV Networks. *IEEE Access* **2019**, *7*, 61786–61796. [CrossRef]
17. Zhao, H.; Wei, J.; Huang, S.; Zhou, L.; Tang, Q. Regular topology formation based on artificial forces for distributed mobile robotic networks. *IEEE Trans. Mob. Comput.* **2019**, *18*, 2415–2429. [CrossRef]
18. Trotta, A.; Di Felice, M.; Montori, F.; Chowdhury, K.R.; Bononi, L. Joint Coverage, Connectivity, and Charging Strategies for Distributed UAV Networks. *IEEE Trans. Robot.* **2018**, *34*, 883–900. [CrossRef]
19. Trotta, A.; Montecchiari, L.; Di Felice, M.; Bononi, L. A GPS-Free Flocking Model for Aerial Mesh Deployments in Disaster-Recovery Scenarios. *IEEE Access* **2020**, *8*, 91558–91573. [CrossRef]
20. Huang, X.; Liu, A.; Zhou, H.; Yu, K.; Wang, W.; Shen, X. FMAC: A Self-Adaptive MAC Protocol for Flocking of Flying Ad Hoc Network. *IEEE Internet Things J.* **2021**, *8*, 610–625. [CrossRef]
21. Hong, L.; Guo, H.; Liu, J.; Zhang, Y. Toward Swarm Coordination: Topology-Aware Inter-UAV Routing Optimization. *IEEE Trans. Veh. Technol.* **2020**, *69*, 10177–10187. [CrossRef]
22. Liu, J.; Wang, Q.; He, C.; AJaffrès-runserc, K.; Xu, Y.; Li, Z.; Xu, Y. QMR:Q-learning based Multi-objective optimization Routing protocol for Flying Ad Hoc Networks. *Comput. Commun.* **2020**, *150*, 304–316. [CrossRef]
23. Arafat, M.Y.; Moh, S. A Q-Learning-Based Topology-Aware Routing Protocol for Flying Ad Hoc Networks. *IEEE Internet Things J.* **2021**, *4662*, 1. [CrossRef]
24. da Costa, L.A.L.F.; Kunst, R.; Pignaton de Freitas, E. Q-FANET: Improved Q-learning based routing protocol for FANETs. *Comput. Netw.* **2021**, *198*, 108379. [CrossRef]
25. Bithas, P.S.; Michailidis, E.T.; Nomikos, N.; Vouyioukas, D.; Kanatas, A.G. A survey on machine-learning techniques for UAV-based communications. *Sensors* **2019**, *19*, 5170. [CrossRef] [PubMed]

26. Mozaffari, M.; Saad, W.; Bennis, M.; Nam, Y.H.; Debbah, M. A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2334–2360. [CrossRef]
27. Shen, G.; Lei, L.; Li, Z.; Cai, S.; Zhang, L.; Cao, P.; Liu, X. Deep Reinforcement Learning for Flocking Motion of Multi-UAV systems: Learn from a Digital Twin. *IEEE Internet Things J.* **2021**, *4662*, 1. [CrossRef]
28. Chen, Q.; Meng, W.; Han, S.; Li, C.; Chen, H.H. Reinforcement Learning-Based Energy-Efficient Data Access for Airborne Users in Civil Aircrafts-Enabled SAGIN. *IEEE Trans. Green Commun. Netw.* **2021**, *5*, 934–949. [CrossRef]
29. Sliwa, B.; Schuler, C.; Patchou, M.; Wietfeld, C. PARRoT: Predictive Ad-hoc Routing Fueled by Reinforcement Learning and Trajectory Knowledge. *IEEE Veh. Technol. Conf.* **2021**, 1–7. [CrossRef]
30. Oubbati, O.S.; Atiquzzaman, M.; Lorenz, P.; Tareque, M.H.; Hossain, M.S. Routing in flying Ad Hoc networks: Survey, constraints, and future challenge perspectives. *IEEE Access* **2019**, *7*, 81057–81105. [CrossRef]
31. Shumeye Lakew, D.; Sa'ad, U.; Dao, N.N.; Na, W.; Cho, S. Routing in Flying Ad Hoc Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1071–1120. [CrossRef]
32. Arafat, M.Y.; Moh, S. Routing protocols for unmanned aerial vehicle networks: A survey. *IEEE Access* **2019**, *7*, 99694–99720. [CrossRef]
33. Leonov, A.V.; Litvinov, G.A.; Korneev, D.A. Simulation and Analysis of Transmission Range Effect on AODV and OLSR Routing Protocols in Flying Ad Hoc Networks (FANETs) formed by Mini-UAVs with Different Node Density. In Proceedings of the 2018 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Minsk, Belarus, 4–5 July 2018. [CrossRef]
34. Garcia-Santiago, A.; Castaneda-Camacho, J.; Guerrero-Castellanos, J.F.; Mino-Aguilar, G. Evaluation of AODV and DSDV routing protocols for a FANET: Further results towards robotic vehicle networks. In Proceedings of the 2018 IEEE 9th Latin American Symposium on Circuits & Systems (LASCAS), Puerto Vallarta, Mexico, 25–28 February 2018; pp. 1–4. [CrossRef]
35. Li, T.; Li, C.; Yang, C.; Shao, J.; Zhang, Y.; Pang, L.; Chang, L.; Yang, L.; Han, Z. A Mean Field Game-Theoretic Cross-Layer Communications. *J. Commun. Netw.* **2022**, *24*, 68–82. [CrossRef]
36. Tan, X.; Zuo, Z.; Su, S.; Guo, X.; Sun, X.; Jiang, D. Performance Analysis of Routing Protocols for UAV Communication Networks. *IEEE Access* **2020**, *8*, 92212–92224. [CrossRef]
37. Bujari, A.; Palazzi, C.E.; Ronzani, D. A Comparison of Stateless Position-based Packet Routing Algorithms for FANETs. *IEEE Trans. Mob. Comput.* **2018**, *17*, 2468–2482. [CrossRef]
38. Oubbati, O.S.; Lakas, A.; Zhou, F.; Güneş, M.; Yagoubi, M.B. A survey on position-based routing protocols for Flying Ad hoc Networks (FANETs). *Veh. Commun.* **2017**, *10*, 29–56. [CrossRef]
39. Jung, W.; Yim, J.; Ko, Y. QGeo: Q-Learning-Based Geographic Ad Hoc Routing Protocol for Unmanned Robotic Networks. *IEEE Commun. Lett.* **2017**, *21*, 2258–2261. [CrossRef]
40. Alam, M.M.; Moh, S. Q-Learning-Based Routing in Flying Ad Hoc Networks: A Survey. In Proceedings of the 10th International Conference on Smart Media and Applications (SMA 2021), Gunsan, Korea, 9–11 September 2021; p. 2021.
41. Ding, R.; Gao, F.; Shen, X.S. 3D UAV Trajectory Design and Frequency Band Allocation for Energy-Efficient and Fair Communication: A Deep Reinforcement Learning Approach. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 7796–7809. [CrossRef]
42. Deb, P.K.; Mukherjee, A.; Misra, S. XiA: Send-it-Anyway Q-Routing for 6G-Enabled UAV-LEO Communications. *IEEE Trans. Netw. Sci. Eng.* **2021**, *8*, 2722–2731. [CrossRef]
43. Jian, X.; Leng, P.; Wang, Y.; Alrashoud, M.; Hossain, M.S. Blockchain-Empowered Trusted Networking for Unmanned Aerial Vehicles in the B5G Era. *IEEE Netw.* **2021**, *35*, 72–77. [CrossRef]
44. Chen, D.; Qi, Q.; Zhuang, Z.; Wang, J.; Liao, J.; Han, Z. Mean Field Deep Reinforcement Learning for Fair and Efficient UAV Control. *IEEE Internet Things J.* **2021**, *8*, 813–828. [CrossRef]
45. Lúís, M.; Oliveira, R.; Bernardo, L.; Garrido, A.; Pinto, P. Joint topology control and routing in ad hoc vehicular networks. In Proceedings of the 2010 European Wireless Conference (EW), Lucca, Italy, 12–15 April 2010; pp. 528–535. [CrossRef]
46. Trotta, A.; Muncuk, U.; Di Felice, M.; Chowdhury, K.R. Tracking Using Unmanned Aerial. *IEEE Veh. Technol. Mag.* **2020**, *15*, 96–103. [CrossRef]
47. Wang, N.; Dai, J.; Ying, J. Research on Consensus of UAV Formation Trajectory Planning Based on Improved Potential Field. In Proceedings of the 2021 40th Chinese Control Conference (CCC), Shanghai, China, 26–28 July 2021; pp. 99–104.
48. Kuiper, E.; Nadjm-Tehrani, S. Mobility models for UAV group reconnaissance applications. In Proceedings of the 2006 International Conference on Wireless and Mobile Communications (ICWMC 06), Bucharest, Romania, 29–31 July 2006; pp. 2–8. [CrossRef]
49. Kieffer, E.; Danoy, G.; Bouvry, P.; Nagih, A. Hybrid mobility model with pheromones for UAV detection task. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2017; pp. 1–8. [CrossRef]
50. Jin, W.; Gu, R.; Ji, Y. Reward Function Learning for Q-learning-Based Geographic Routing Protocol. *IEEE Commun. Lett.* **2019**, *23*, 1236–1239. [CrossRef]
51. He, C.; Wang, Q.; Xu, Y.; Liu, J.; Xu, Y. A q-learning based cross-layer transmission protocol for MANETs. In Proceedings of the 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), Shenyang, China, 21–23 October 2019; pp. 580–585. [CrossRef]

52. Javaid, N.; Bibi, A.; Djouani, K. Interference and bandwidth adjusted ETX in wireless multi-hop networks. In Proceedings of the 2010 IEEE Globecom Workshops, Miami, FL, USA, 6–10 December 2010; pp. 1638–1643. [CrossRef]
53. Faganello, L.R.; Kunst, R.; Both, C.B.; Granville, L.Z.; Rochol, J. Improving reinforcement leatwinrning algorithms for dynamic spectrum allocation in cognitive sensor networks. In Proceedings of the 2013 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 7–10 April 2013; pp. 35–40. [CrossRef]
54. Behnke, D.; Bök, P.B.; Wietfeld, C. UAV-based connectivity maintenance for borderline detection. In Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), Dresden, Germany, 2–5 June 2013; pp. 2–7. [CrossRef]
55. Sliwa, B.; Patchou, M.; Wietfeld, C. Lightweight simulation of hybrid aerial- And ground-based vehicular communication networks. In Proceedings of the 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), Honolulu, HI, USA, 22–25 September 2019; pp. 1–7. [CrossRef]
56. Rovira-Sugranes, A.; Afghah, F.; Qu, J.; Razi, A. Fully-echoed Q-routing with Simulated Annealing Inference for Flying Adhoc Networks. *IEEE Trans. Netw. Sci. Eng.* **2021**, *8*, 2223–2234. [CrossRef]
57. Yang, Q.; Jang, S.J.; Yoo, S.J. Q-Learning-Based Fuzzy Logic for Multi-objective Routing Algorithm in Flying Ad Hoc Networks. *Wirel. Pers. Commun.* **2020**, *113*, 115–138. [CrossRef]
58. Zhang, M.; Dong, C.; Yang, P.; Tao, T.; Wu, Q.; Quek, T.Q.S. Adaptive Routing Design for Flying Ad Hoc Networks. *IEEE Commun. Lett.* **2022**, *14*, 1. [CrossRef]
59. Oubbati, O.S.; Lakas, A.; Lorenz, P.; Atiquzzaman, M.; Jamalipour, A. Leveraging communicating UAVs for emergency vehicle guidance in Urban Areas. *IEEE Trans. Emerg. Top. Comput.* **2021**, *9*, 1070–1082. [CrossRef]
60. Li, L.; Cheng, Q.; Xue, K.; Yang, C.; Han, Z. Downlink Transmit Power Control in Ultra-Dense UAV Network Based on Mean Field Game and Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2020**, *69*, 15594–15605. [CrossRef]
61. Xiao, Z.; Zhu, L.; Liu, Y.; Yi, P.; Zhang, R.; Xia, X.G.; Schober, R. A Survey on Millimeter-Wave Beamforming Enabled UAV Communications and Networking. *IEEE Commun. Surv. Tutor.* **2021**, *24*, 557–610. [CrossRef]
62. Khawaja, W.; Guvenc, I.; Matolak, D.W.; Fiebig, U.C.; Schneckenburger, N. A Survey of Air-to-Ground Propagation Channel Modeling for Unmanned Aerial Vehicles. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2361–2391. [CrossRef]
63. Liu, Y.; Wang, Y.; Wang, J.; Shen, Y. Distributed 3D Relative Localization of UAVs. *IEEE Trans. Veh. Technol.* **2020**, *69*, 11756–11770. [CrossRef]
64. Nemer, I.A.; Sheltami, T.R.; Mahmoud, A.S. A game theoretic approach of deployment a multiple UAVs for optimal coverage. *Transp. Res. Part A Policy Pract.* **2020**, *140*, 215–230. [CrossRef]
65. Mason, F.; Capuzzo, M.; Magrin, D.; Chiariotti, F.; Zanella, A.; Zorzi, M. Remote Tracking of UAV Swarms via 3D Mobility Models and LoRaWAN Communications. *IEEE Trans. Wirel. Commun.* **2021**, 1–16. [CrossRef]
66. Kia, S.S.; Rounds, S.; Martınez, S. Cooperative Localization for Mobile Agents. *IEEE Control Syst. Mag.* **2016**, *36*, 86–101.
67. Wu, Q.; Zhang, M.; Dong, C.; Feng, Y.; Yuan, Y.; Feng, S.; Quek, T.Q.S. Routing protocol for heterogeneous FANETs with mobility prediction. *China Commun.* **2022**, *19*, 186–201. [CrossRef]
68. Seyyedabbasi, A.; Aliyev, R.; Kiani, F.; Gulle, M.U.; Basyildiz, H.; Shah, M.A. Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems. *Knowl. Based Syst.* **2021**, *223*, 107044. [CrossRef]
69. Jiang, X.; Sheng, M.; Zhao, N.; Xing, C.; Lu, W.; Wang, X. Green UAV communications for 6G: A survey. *Chin. J. Aeronaut.* **2021**. [CrossRef]
70. Jia, Z.; Sheng, M.; Li, J.; Han, Z. Towards Data Collection and Transmission in 6G Space-Air-Ground Integrated Networks: Cooperative HAP and LEO Satellite Schemes. *IEEE Internet Things J.* **2021**, *4662*, 1. [CrossRef]